

**Seasonal surface urban energy balance and wintertime stability simulated using  
three land-surface models in the high-latitude city Helsinki**

P. Karsisto<sup>1,2</sup>, C. Fortelius<sup>2</sup>, M. Demuzere<sup>3</sup>, C. S. B. Grimmond<sup>4</sup>, K. W. Oleson<sup>5</sup>, R.  
Kouznetsov<sup>2,7</sup>, V. Masson<sup>6</sup> and L. Järvi<sup>1,\*</sup>

<sup>1</sup>Department of Physics, University of Helsinki, Finland

<sup>2</sup>Finnish Meteorological Institute, Finland

<sup>3</sup>KU Leuven, Department of Earth and Environmental Sciences, Leuven, Belgium

<sup>4</sup>Department of Meteorology, University of Reading, UK

<sup>5</sup>National Center for Atmospheric Research, Boulder, USA

<sup>6</sup>CNRM –GAME, Météo-France, CNRS, Toulouse and Grenoble, France

<sup>7</sup>Obukhov Institute for Atmospheric Physics, Moscow, Russia

\*Corresponding author: L. Järvi, leena.jarvi@helsinki.fi

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/qj.2659

## Abstract

The performance of three urban land surface models, run in offline mode, with their default external parameters, is evaluated for two distinctly different sites in Helsinki: Torni and Kumpula. The former is a dense city centre site with 22% vegetation, while the latter is a suburban site with over 50% vegetation. At both locations the models are compared against sensible and latent heat fluxes measured using the eddy covariance technique, along with snow depth observations. The cold climate experienced by the city causes strong seasonal variations that include snow cover and stable atmospheric conditions.

Most of the time the three models are able to account for the differences between the study areas as well as the seasonal and diurnal variability of the energy balance components. However, the performances are not systematic across the modelled components, season and surface type. The net all-wave radiation is well simulated, with the greatest uncertainties related to snowmelt timing, when the fraction of snow cover has a key role, particularly in determining the surface albedo. For the turbulent fluxes, more variation between the models is seen which can partly be explained by the different methods in their calculation and partly by surface parameter values. For the sensible heat flux, simulation of wintertime values was the main problem, which also leads to issues in predicting near-surface stabilities particularly at the dense city centre site. All models have the most difficulties in simulating latent heat flux. This study particularly emphasizes that improvements are needed in the parameterization of anthropogenic heat flux and thermal parameters in winter, snow cover in spring and evapotranspiration in order to improve the surface energy balance modelling in cold climate cities.

Keywords: CLM, eddy covariance, high-latitude, stability, SUEWS, surface energy balance, SURFEX, urban

## 1. Introduction

With the increasing number and density of people living in urban areas (World Bank, 2014) it is crucial to be able to forecast the atmospheric conditions in these areas for a wide range of reasons. This includes for extreme conditions (e.g. heatwaves, intense precipitation) but also day-to-day variations in air quality, plus for long term planning and design (McMichael *et al.*, 2006; IPCC, 2013). Enhanced urban effects, combined with large exposure to these adverse effects, require appropriate actions to be taken. In particular, tools capable and appropriate to examine and predict high latitudes atmospheric condition are needed, as the most dramatic climate change is expected in these areas (IPCC, 2013). High-latitude areas, in comparison to mid- or tropical latitude areas, are characterized by strong seasonal variability in meteorological conditions, particularly in solar radiation and air temperature, impacting vegetation phenology and heating need. In addition, snow on the ground is a common sight in winter and spring altering the surface conditions by increasing surface albedo and modifying surface water availability and heat storage (Semádeni-Davies and Bengtsson, 1998). Additionally, the boundary layer can remain very stable in urban areas during winter restricting pollutant dispersion, and creating the potential for hazardous conditions. Therefore, prediction of atmospheric stability is important for high-latitude communities. However, the stable conditions commonly cause challenges for weather prediction and air quality models (Holtslag *et al.*, 2013).

The description of different surfaces in atmospheric models is provided by land surface models (LSMs) that parameterize the surface-atmosphere exchanges. Many urban LSMs, with varying degree of complexity, were evaluated by Grimmond *et al.* (2011). They demonstrated that the models do not perform well across all components indicating lack of understanding of the urban processes, in particular evapotranspiration. Thus, a better description of the factors affecting the energy partitioning into sensible ( $Q_H$ ) and latent heat fluxes ( $Q_E$ ) in urban areas are needed. In urban areas, the energy balance equation can be written:

$$Q^* + Q_F = Q_E + Q_H + \Delta Q_S + \Delta Q_A, \quad (\text{W m}^{-2}) \quad (1)$$

where  $Q^*$  is the net all-wave radiation,  $Q_F$  is the anthropogenic heat flux, and  $\Delta Q_S$  is the net storage heat flux including both soil heat flux and heat storage and release of urban fabric.  $\Delta Q_A$  is net heat advection and typically in land surface models micro-scale (or sub-grid scale) advection is included within the parameterizations of the individual energy balance terms. In Eq. (1), only  $Q^*$  and the turbulent fluxes can be easily measured. The other components require detailed measurements combined with models to be estimated.

Relatively little attention has been given to the performance of LSMs in cold climate or high-latitude cities (Lemonsu *et al.*, 2010; Leroyer *et al.*, 2010; Järvi *et al.*, 2014). Thus the purpose of this study is to evaluate and compare the performance of three urban LSMs for two areas with different urban land covers in Helsinki. The models are run in their default mode in order to understand how the model physics and default parameters, commonly derived and successfully used in mid-latitude cities, perform in a high-latitude city. The models used are the Community Land Model

(CLM, Lawrence *et al.* (2011)), the Surface Urban Energy and Water Balance Scheme (SUEWS, Järvi *et al.* (2011) and Surface Externalisée (SURFEX, Masson *et al.* (2013)). The evaluation dataset consists of: net all-wave radiation, turbulent sensible and latent heat fluxes measured using the eddy covariance (EC) technique, and snow depth. The effects of the model performances on energy partitioning and predicting near-surface stability are examined. Following the study site and measurement setup (Section 2), the model descriptions (Section 3) are given. The specific model setups and analysis methods are provided in Section 4. Finally, the results are shown and discussed (Section 5) and conclusions drawn (Section 6).

## 2. Study sites

The focus of this study is two sites within the capital city of Finland, Helsinki (Figure 1a, Table 1). This northern European city (latitude 60°N) has approximately 616 000 inhabitants (Tilastokeskus, 2014). Despite the clearly reduced wintertime solar radiation (Figure 1b), the Atlantic Ocean and the Baltic Sea create milder winters than typically observed at this latitude. Helsinki's heating is commonly centralized and the buildings are designed to retain heat.

The intra-annual variability of daytime mean between 1000-1400 local time (UTC+2) of incoming shortwave radiation ( $K\downarrow$ ), and daily mean air temperature ( $T_{air}$ ), relative humidity (RH), and incoming longwave radiation ( $L\downarrow$ ), plus daily total precipitation ( $P$ ) are shown in Figure 1b for 2012. In winter, daytime mean  $K\downarrow$  stays below 100 W m<sup>-2</sup> whereas in summer it reaches 750 W m<sup>-2</sup>. The reduced solar radiation has implications for both the available energy in the system and vegetation phenology;

both of which are crucial for surface energy balance (Eq. 1) behaviour. In February, the daily mean  $T_{air}$  is  $-20^{\circ}\text{C}$  whereas in July it is over  $+20^{\circ}\text{C}$ . As winter  $T_{air}$  stays for most of the time below freezing most  $P$  occurs as snowfall. Thus, the significant role snow plays in the winter urban energy balance has to be taken into account. RH varies between 20% and 100% with the lowest values in spring and early summer.  $L\downarrow$  follows closely the annual behaviour of  $T_{air}$  with a range of 200 to  $350\text{ W m}^{-2}$ .  $P$  is distributed through the year, with slightly higher rates observed in autumn in 2012.

The study areas (Figure 1a) are based on the approximate turbulent flux footprints of SMEAR (*Station for Measuring Ecosystem- Atmosphere-Relations*) III observations. The first site, suburban Kumpula, is located 4 km north-east from the Helsinki city centre and can be characterized as a local climate zone (LCZ) 6 (Stewart and Oke, 2012). The 31 m high measurement mast is surrounded with three distinct surface cover sectors (Vesala *et al.*, 2008). To the north ( $320\text{--}40^{\circ}$ , referred to here as Ku1) in the foreground are the University of Helsinki campus buildings and further away suburban apartments and generally low building heights with small gardens. To the east ( $40\text{--}180^{\circ}$ , Ku2) is a large road heading to Helsinki city centre (closest distance of 150 m) with mixed broadleaf forest in the foreground. Finally, to the west ( $180\text{--}320^{\circ}$ , Ku3) a vegetated area consists of allotment gardens and the University Botanical Garden. As the anthropogenic activities, including heating, traffic and irrigation, as well as vegetation types vary between the three sectors – the road, vegetation and urban – these areas are modelled separately.

The second site (Hotel Tornio), in the Helsinki city centre (Nordbo *et al.*, 2013), is highly built-up with only slightly more than 20% covered with vegetation. There are a

few small parks with shrubs and trees plus some street canyons with trees. Most of the buildings are 4-5 storey buildings with a mean height of 18 m. This site is classified as LCZ 2. As the city centre is located on a peninsula, the Baltic Sea is near the measurement site in three wind directions.

## 2.1 Measurements

Both sites have EC systems that allow evaluation of model performances for the turbulent heat and momentum fluxes. The EC technique directly measures the heat and moisture carried by the turbulence in the atmosphere (Aubinet *et al.*, 2012). With simultaneous measurements of vertical wind speed and the scalar of interest (e.g. temperature), the vertical exchange is calculated as a covariance between these two, from which the fluxes can be calculated.

At both sites wind components and sonic temperature are measured with an ultrasonic anemometer (USA-1, Metek GmbH, Germany). The water vapour mixing ratio is measured at Kumpula with a closed-path infrared gas analyser (LI-7000, LI-COR, Lincoln, Nebraska, USA) and at Hotel Tornio using an enclosed path analyser (LI-7200, LI-COR). The 10 Hz measurements were analysed using commonly accepted methods (Nordbo *et al.*, 2012). The EC measurements are carried out at a sufficient height above the surrounding buildings at both sites (31 m Kumpula, 60 m Tornio).

Meteorological data are needed both for model forcing and evaluation. Most of the measurements are made at Kumpula, where the measurements include tower based  $T_{air}$  (platinum resistant thermometer, Pt-100, "in-house"), wind speed (Thies Clima

2.1x, Goettingen, Germany) and incoming and outgoing short- and long-wave radiation (CNR1, Kipp&Zonen, Delft, Netherlands) at 31 m above ground level (agl). Air pressure (DPA500, Vaisala Oyj, Vantaa, Finland), relative humidity (HMP243, Vaisala Oyj), and precipitation (rain gauge, Pluvio2, Ott Messtechnik GmbH, Germany) are measured from the roof of a nearby building at 24 m agl. Downtown,  $T_{air}$  (HMP45D, Vaisala Oyj) and outgoing short- and long-wave radiation (CNR4, Kipp&Zonen) are measured 550 m southeast of the Tornio site at a height of 53 m. Snow depths are measured near the Kumpula mast and in a park in city centre by the Finnish Meteorological Institute.

### 3. Model description

Each of the LSM can be coupled to an atmospheric model (e.g. for numerical weather prediction or climate simulation), or as in this study, used as a standalone model. Given the focus on anthropogenic heat and snow cover in urban areas, these parameterizations are discussed below (see also Table 2).

#### 3.1 CLM

Version 4 of the Community Land Model (CLM4, Lawrence *et al.* (2011)), the land surface scheme embedded in the Community Earth System Model (CESM, Hurrell *et al.* (2013)), is used here. CLM4 uses a nested sub-grid hierarchy in which each grid cell (or tower footprint) can have up to five land units: wetlands, glaciers, vegetation, lakes and urban. The vegetated fraction is modelled with the vegetation canopy/soil model in CLM and the urban fraction, or urban land unit, with the Community Land



Model Urban (CLMU, Oleson *et al.* (2008a,b)). CLMU follows the concept of Oke (1987) in which the considerable complexity of an urban environment is reduced to a single-layer urban canyon that consists of five columns: roof, sunlit and shaded wall, impervious and pervious road. The walls are hydrologically inactive, while liquid and solid precipitation can be intercepted, stored and evaporated from the roof and canyon floor (both impervious and pervious road). The advantage of such a modelling framework is that the pervious fraction is an integral part of the urban canyon and will thus interact with urban canyon air properties such as humidity and temperature (Oleson *et al.*, 2008b; Demuzere *et al.*, 2013). The required model input of  $K_d$  is partitioned into direct and diffuse radiation based on empirical factors derived from analysis of one year of hourly Community Atmosphere Model (CAM) output (Oleson *et al.*, 2010a). For full technical descriptions of CLMU and CLM, see Oleson *et al.* (2010a, b).

CLMU provides three options to calculate  $Q_F$ : 1) assume that it is negligible, 2) limit the internal building temperature to a specified comfort range with no waste heat released into the urban canopy layer, or 3) with waste heat released into the urban canopy layer (see also Demuzere *et al.* (2013)). For Helsinki, option 3 is used with an internal minimum building temperature ( $T_{min,building}$ ) of 19°C (heating required) but no maximum internal building temperature is used as in Jackson *et al.* (2010) this is set sufficiently high for this high-latitude region such that air conditioning is never required. Here,  $T_{min,building}$  is used as a lower boundary condition in the solution of heat conductivity in roof and walls. The total waste heat flux into the urban canopy layer is controlled by factors describing the efficiencies of space heating/air conditioning systems and the conversion of primary to end use energy (see eq. 4.55 in

Oleson *et al.* (2010a)). For this study, the factor for waste heat from space heating is set to 0.2 based on analysis by Sivak (2013).

In CLMU, a snowpack (represented by up to 5 snow layers) that can form on the active surfaces (roof and road fractions) influences processes in the energy and hydrological cycles. The albedo and emissivity of each urban surface are a weighted combination of snow-free and snow albedos. A fractional snow cover ( $f_s$ ) is defined for roof and canyon floor surfaces as  $f_s = s_d/0.05 \leq 1$  with  $s_d$  being the snow depth in metres. From a hydrological perspective, the roof and impervious road are able to store snow, while the pervious road fraction includes snow accumulation and melt and water transfer between snow layers besides other hydrological processes such as infiltration, surface runoff, subsurface drainage, and redistribution within the soil column.

### 3.2 SUEWS

The Surface Urban Energy and Water balance Scheme (SUEWS, Järvi *et al.* (2011)) simulates the urban water and energy balances at the local or neighbourhood scale. Here version V2014b is used. It requires basic meteorological variables and information about the surface cover of the study area including surface cover fractions, tree and building heights, and number of people. The model determines the rates of evaporation/interception at a 5 to 60 minute time step for seven surfaces types (paved, buildings, evergreen trees/shrubs, deciduous trees/shrubs, grass, unmanaged and water). Below each surface, except water, there is a single soil layer. In SUEWS, the surface types are not separate tiles but rather water can flow between the surface

types both above and below surface. Integrated evapotranspiration for the whole study area is calculated using the modified Penman-Monteith equation (Grimmond and Oke, 1991).

SUEWS can simulate all components of  $Q^*$  using the net all-wave radiation parameterization scheme (NARP, Offerle *et al.*, 2003; Loridan *et al.*, 2011) except  $K\downarrow$  that is a required model input. However, in this study the provided  $L\downarrow$  is used to be comparable in the model evaluation. Storage heat flux is simulated for each surface type using the Objective Hysteresis Model (OHM, Grimmond *et al.* (1991)). Calculation of  $Q_F$  is based on heating and cooling degree-days assuming a base air temperature of 18.2°C above which cooling and below heating of the buildings is assumed.  $Q_F$  is also affected by the population density of the study area.

A single layer snowpack (also freezing of water) can develop on all surface types. The snow energy and mass balances are calculated at each time step (Järvi *et al.*, 2014). In the model, snowfall occurs and snowpack develops and existing water on surfaces freezes when  $T_{air}$  is below 0°C. Snowmelt is calculated using a temperature degree-day method based on  $T_{air}$  and solar radiation and the melted water will stay in the snowpack until the water holding capacity, for each surface type, is exceeded after which runoff occurs. Snow heat storage is calculated with OHM. The temporal evolution of snow density and albedo use snow aging functions.  $f_s$  on ground is calculated from depletion curves, where  $f_s$  is a different function of  $W_s/W_{s,max}$  for vegetation, paved and building surfaces.  $W_s$  is snow water equivalent and  $W_{s,max}$  is the maximum value, both defined by the user. The energy and mass balances at the snow and snow-free surfaces are calculated separately and the model output is calculated as

a weighted average based on their surface fractions.

SUEWS is the only model in this study where leaf area index (LAI) is allowed to dynamically vary between minimum and maximum value of each vegetation type based on growing and senescence degree days (Järvi *et al.*, 2011). Thus, changes in LAI can vary from year to year based on  $T_{air}$ .

### 3.3 SURFEX

SURFEX (Masson *et al.*, 2013) is a system of models able to calculate the exchange of sensible and latent heat, momentum, carbon dioxide and other chemical species, as well as various particles, between the atmosphere and several types of surfaces. Here version 7.3 is used. Four main types of surfaces can be treated: sea surfaces, inland waters, a large variety of vegetated land surfaces, and urban environments. Separate models, often developed and evaluated independently of SURFEX, are used to describe the surface fluxes and evolution inside each main surface type. Modelling of the snow pack and hydrological processes generating surface and bottom run-off are also included.

Vegetation is treated by the Interaction Soil–Biosphere–Atmosphere model (ISBA) (Masson *et al.*, 2013), which can be configured to varying degrees of complexity (Noilhan and Planton, 1989). In the present study, the soil is modelled by a three-layer force-restore formulation distinguishing between the root zone and the subroot zone, and also taking into account the freezing and melting of water in the soil (Boone *et al.*, 2000). Vegetation processes are described by the original ISBA

evapotranspiration model using an externally imposed LAI. The snow pack is treated by the bulk snow model (Douville *et al.*, 1995) with three prognostic variables ( $W_s$ , snow density, and snow albedo).

The urban surface uses the Town Energy Balance Model (TEB, Masson (2000)), to solve separate energy budgets for roofs, walls, and road surfaces in a homogeneous isotropic array of street canyons. Although TEB has vegetation, parks and gardens inside the TEB canyon itself (Lemonsu *et al.*, 2012), here the vegetated areas are treated separately without direct interaction with the built surface (e.g. buildings shadowing). Snow is treated as a single layer, and on the vegetation it is simulated by ISBA, and on roofs and roads by TEB. The snow temperature, albedo and density evolve governed by the energy budgets interacting with the atmosphere and the underlying surface. Prior evaluation of TEB was for winter conditions in Montreal, Canada (Lemonsu *et al.*, 2010), but it did not account for seasonal differences in model performance.

## 4. Methods

### 4.1 Model runs

The model comparisons for the two sites use the same hourly meteorological forcing data from Kumpula ( $K\downarrow$ ,  $L\downarrow$ ,  $P$ , wind speed, pressure and humidity) except for  $T_{air}$ , which was measured both in Kumpula and in Helsinki city centre. Wind speed measured at 31 m above ground level (a.g.l.) at Kumpula was not scaled to the 60 m (a.g.l.) measurement height at Tornio due to possible source of uncertainty originating

from differences in surface roughness and topography. The 18-month forcing data were gap-filled (2.4 % of  $K\downarrow$ ,  $L\downarrow$ ,  $T_{air}$ , wind speed; 0.7 % of  $P$ ; 2.3 % of pressure and humidity) with observations from other instruments (if available) and via linear interpolation (Järvi *et al.*, 2012).

All models use a set of same surface parameters (Table 1) calculated for 1 km radius circles centred on both measurement masts (Figure 1a) to approximate the EC flux source areas (Nordbo *et al.*, 2013). These circles are chosen over the actual source areas, as no footprint model yet exists to correctly simulate source areas over heterogeneous urban surfaces. The Kumpula site heterogeneity is addressed by modelling the three sectors (Ku1, Ku2 and Ku3) using separate parameters and creating one time series based on the prevailing wind direction. In Helsinki, the most common wind direction is southwest, which corresponds to the most vegetated sector Ku3. In the case of  $Q^*$ , only model output from Ku3 is considered as it best describes the source area of the radiation measurements. These surface parameters include surface cover fractions, building and tree heights and the initial conditions for the run. As the three models have different approaches to the vegetation subclasses these fractions are not identical (Table 1).

For the other model parameters, CLM uses the Jackson *et al.* (2010) database, SURFEX the ECOCLIMAP (Masson *et al.*, 2003; Faroux *et al.*, 2013) and SUEWS its own values (Järvi *et al.*, 2011). The chosen parameters are default parameters most suitable for high latitude city without any adjustments based on the observations. These include albedos, emissivities, roughness parameters, parameters for plant function types, geometrical and thermal parameters for the buildings and roads, and

parameters for anthropogenic heating. As the snow model in SUEWS has partly been developed using observations from the Kumpula site (Järvi *et al.*, 2014), its performance is not purely independent at the site unlike SURFEX and CLM. Table 2 lists some of the used parameter values important for cold climate and high latitude cities.

The first 6 months of the 18-month model period (1 July 2011 to 31 December 2012) are used to spin-up the models. A full year (2012) is used for the model evaluation.

#### 4.2 Thermal seasons

Given the strong seasonal variability in meteorological conditions in Helsinki, analysis is based on thermally defined seasons (rather than months) in this study calculated from the 7-day running mean of  $T_{air}$ . If the running mean is below 0°C or above 10°C the season is winter or summer, respectively. Between these limits are spring and autumn. Seasons change when the temperature criterion is fulfilled for 3 days in a row. Thus, thermal seasons vary by length each year. In 2012, summer was the longest season (39% of the days), extending from mid-May to 6<sup>th</sup> October (Fig. 1b). The two winter periods (9 January – 14 March, 2 – 31 December) resulted in a total of 95 days (26%), whereas autumn (1 – 8 January, 7 October – 1 December) and spring (15 March – 14 May) covered only 19% and 16% of the days, respectively.

#### 4.3 Goodness of models

The model performance is evaluated using common tools, including the Taylor

diagram (Taylor, 2001; Grimmond *et al.*, 2011) and related statistical parameters of root mean square error (RMSE,  $\text{W m}^{-2}$ ), Pearson correlation coefficient ( $r$ ), standard deviation (STD,  $\text{W m}^{-2}$ ) and mean bias error (MBE,  $\text{W m}^{-2}$ ) (Grimmond *et al.*, 2010).

The uncertainties of the EC measurements are taken into account by calculating the absolute difference  $|F_{obs} - F_{mod}|$  between observation ( $F_{obs}$ ) and model output ( $F_{mod}$ ) (Best and Grimmond, 2013). For forested areas, Hollinger and Richardson (2005) reported flux uncertainties for EC based sensible ( $\delta_H$ ) and latent heat flux ( $\delta_E$ ) be at least  $10 \text{ W m}^{-2}$  and increase linearly as a function of the absolute value of the flux:

$$\delta_H = 10 + 0.22|Q_H|, \quad (\text{W m}^{-2}) \quad (2)$$

$$\delta_E = 10 + 0.32|Q_E|. \quad (\text{W m}^{-2}) \quad (3)$$

Their study was conducted in a relatively homogeneous setting, so these uncertainties are likely to be lower limits for heterogeneous urban environments.

Equations 2 and 3 define the acceptable absolute difference between the model output and observed turbulent fluxes: if the absolute difference is within the uncertainty of the EC measurements, the model output is considered to be good ( $(|F_{obs} - F_{mod}| \leq \delta)$ ). Furthermore the normalized acceptable deviation ( $D$ ) is the occurrence of acceptable deviations, defined as those times the mean absolute difference is within the measurement uncertainty divided by the number of observations ( $N_{obs}$ ):

$$D = \frac{\sum(|F_{obs} - F_{mod}| \leq \delta)}{N_{obs}}. \quad (4)$$

$D$  is zero when none of  $N_{obs}$  is within the flux uncertainties and one when all  $N_{obs}$  are within the flux uncertainties.

## 5. Results



## 5.1 Snow cover

Snow cover is one of the most crucial variables to be simulated correctly in cold climate cities as it affects surface properties and surface water availability. This applies both to the amount of snow as well as its areal fraction, which are tied together via depletion curves (Table 2). In Figure 2, the modelled snow depth ( $s_d$ ) and snow surface fraction at the vegetated surface ( $f_{s,veg}$ ) are plotted for the two sites together with the observed  $s_d$ . Unfortunately, no measured  $f_s$  from any of the surfaces is available for the sites. For evaluating the modelled snow data, the most appropriate data matching best with the location of the snow observations are shown: grass surface fraction from SUEWS and grass and trees from SURFEX and CLM simulated at the vegetated Ku3 sector to match the open field observed  $s_d$ .

At Tornio, CLM simulates  $s_d$  well and snow pack develops and melts on the correct day (Figure 2a). SUEWS and SURFEX reproduce the snow pack development but the melt is too slow. The complete snowmelt is 10 and 17 days later than the observations, respectively. Some of the snowmelt timing differences may relate to the observational representativeness. The models give the snow properties for the whole study area, whereas the single point observations in an open field are not necessarily representative of the whole area. Melt especially in the shadows of trees and buildings and in snow piles can be delayed when compared to the open field. However,  $f_{s,veg}$  with SUEWS is 60% and with SURFEX 90% after the observed snow has melted (Figure 2b), and it is unlikely that so much snow would still be on the ground. SUEWS and SURFEX simulate more similar  $f_{s,veg}$  than CLM, with slightly smaller fractions obtained from SUEWS. Most of the time CLM gives smaller  $f_{s,veg}$  than the

other two models. The snowfall events are visible in CLM and SUEWS as increasing peaks, whereas in SURFEX the snowfall episodes do not instantaneously affect  $f_{s,veg}$  as the fraction is calculated based on the absolute snow depth and not the snowfall events.

The models simulate well the accumulation of snow at Kumpula, but again differences appear during the snow-melting period (Figure 2c). SUEWS simulates complete snow melt one day after and SURFEX one day before the observations, whereas CLM melts snow 8 days too early. Also at Kumpula the largest differences in  $f_{s,veg}$  are seen during the melting period (Figure 2d). In CLM and SURFEX, the differences in  $f_{s,veg}$  between the two sites are due to the different fraction of grass and trees at the two sites, as the models do not separate these similarly to SUEWS. The differences both in  $s_d$  and  $f_{s,veg}$  result in differences in model performances, when simulating the energy balance components. These results are further examined in the following sections.

## 5.2 Net all-wave radiation

The diurnal cycles of the modelled and measured  $Q^*$  are similar both at Tornä and Kumpula (Figure 3). All models simulate  $Q^*$  well with  $r$  above 0.97 and RMSEs between 5 and 44 W m<sup>-2</sup> (Table 3) with minor seasonal variance. The models tend to underestimate the daytime  $Q^*$  at Tornä, with the exception of SUEWS in winter and SUEWS and SURFEX in autumn. All models systematically overestimate the outgoing shortwave radiation ( $K \uparrow$ ) at Tornä indicating too high surface albedo in the models (Table 3, Figure S1). However, bias error cancellation (Shaffer *et al.*, 2015)

caused by overestimated  $K\hat{\uparrow}$  and underestimated outgoing longwave radiation ( $L\hat{\uparrow}$ ) improves the winter  $Q^*$  by SUEWS (RMSE = 8.8 W m<sup>-2</sup>) and spring by SUEWS and SURFEX (RMSE = 25.1 and 44.0 W m<sup>-2</sup>) (Table 3, Figure S2). The modelled nocturnal  $Q^*$  follow the observations, but CLM slightly underestimates  $Q^*$  in winter and autumn and SURFEX overestimates in spring and summer.

At Kumpula, more deviations between the models are seen. Part of this can be related to the heterogeneous surface cover, which causes the radiation footprint to differ from the simulated Ku3. SUEWS underestimates the daytime  $K\hat{\uparrow}$  in winter and spring resulting in overestimation of  $Q^*$ . Similarly CLM underestimates daytime  $K\hat{\uparrow}$ , but this is compensated by overestimation of  $L\hat{\uparrow}$  resulting in well-simulated  $Q^*$  (RMSE = 6.8 – 16.3 W m<sup>-2</sup>). SURFEX underestimates  $K\hat{\uparrow}$  in winter, but this is compensated by overestimation of  $L\hat{\uparrow}$  resulting RMSE = 10.2 W m<sup>-2</sup> for  $Q^*$ . In summer, all models simulate  $K\hat{\uparrow}$  correctly (RMSE = 4.4 – 10.2 W m<sup>-2</sup>), but overestimate daytime  $L\hat{\uparrow}$  and underestimate nocturnal  $L\hat{\uparrow}$ . All models perform well in autumn, when the RMSEs vary between 5 – 7 W m<sup>-2</sup>.

Spring is when the models have most difficulty to reproduce the observed diurnal cycle of  $Q^*$  at both sites. This period has the highest RMSEs with a range from 16 W m<sup>-2</sup> (CLM at Kumpula) to 44 W m<sup>-2</sup> (SURFEX at Tornio). Spring is when the surface characteristics vary the most due to both snowmelt and start of the leaf-on period. Looking in more detail the model performances during this period show that the greatest model uncertainties are related to the snow melt period rather than increase in LAI. The models particularly have problems with the late snowfall event in early April (Figure 2a,c) when 4-day RMSEs (not shown) reach 26, 60 and 35 W m<sup>-2</sup> at

Torni and 32, 30 and 27 W m<sup>-2</sup> at Kumpula, using CLM, SUEWS and SURFEX, respectively. Simulating  $f_s$  can be difficult during spring affecting particularly the weighted  $K \uparrow$  and furthermore  $Q^*$  of snow covered and snow free surfaces.

Interestingly, the over prediction of snow cover by SUEWS and SURFEX on the vegetated surface at Torni does not decrease the model performances when simulating  $Q^*$ . This can be explained by the small fraction of vegetated surfaces around the Torni site. On built surfaces, the modelled snow melts earlier than on the vegetated surfaces (not shown) but unfortunately we do not have snow observations to verify if the timing is correct.

### 5.3 Sensible heat flux

Whereas the LSMs more or less agree upon  $Q^*$ , they differ more in simulating  $Q_H$  (Figure 4, Table 4), in agreement with previous model comparison studies (Grimmond *et al.*, 2011). In winter, the observed daytime  $Q_H$  reaches 120 W m<sup>-2</sup> at Torni, whereas at Kumpula  $Q_H$  is less than half of this. The difference is partly explained by the increased anthropogenic heat emissions during the heating period and partly by the release of heat storage to the atmosphere. The larger nocturnal release of storage heat at Torni compared to Kumpula is particularly apparent in autumn.

CLM simulates the wintertime  $Q_H$  best at Kumpula (RMSE = 16 W m<sup>-2</sup>), but overestimates  $Q_H$  at Torni (RMSE = 45 W m<sup>-2</sup>) particularly at night. Although SUEWS is able to predict the difference between the two sites, it underestimates nocturnal  $Q_H$  at Torni and overestimates daytime  $Q_H$  at Kumpula (RMSE of

respectively  $29 \text{ W m}^{-2}$  and  $19 \text{ W m}^{-2}$ ). SURFEX underestimates  $Q_H$  at both Tornø and Kumpula (RMSE =  $36 \text{ W m}^{-2}$  and  $20 \text{ W m}^{-2}$ , respectively) and is not able to predict the difference between the two sites in wintertime. As CLM's urban module (to a large extent) follows the parameterisations of SURFEX (Oleson *et al.*, 2008b; Demuzere *et al.*, 2013), it is surprising to see such large differences between the two models, especially in winter for Tornø. To examine this, additional CLM simulations were performed using the same thermal (roof and wall thickness, heat conductivity and volumetric heat capacity) and radiative (albedo and emissivity for roof, road and wall) properties as used in the SURFEX simulations. As these caused  $Q_H$  to decrease by  $100 \text{ W m}^{-2}$  (not shown) and thus more similar to the SURFEX results, it suggests the Jackson *et al.* (2010) high-latitude parameters should be revisited. The radiative parameters affected particularly the outgoing shortwave radiation, but most of the reduction resulted from the thermal parameters. Particularly noteworthy is the lack of insulation in the Jackson *et al.* (2010) roof and wall properties for this high-latitude region. This means that more space heating is needed to keep the internal building temperature from falling below the prescribed minimum comfort level. The increase in space heating is manifested as a strong sensible heat from the interior of the building to the exterior of the roofs and walls and the urban canopy air. When the SURFEX roof/wall thermal parameters are used (increased insulation), the space heating required and thus the sensible heat decreases. This agrees well with previous findings of Oleson *et al.* (2008a) who found that  $Q_H$  in CLM is most sensitive to the morphological and thermal parameters.

Although in other seasons, the observed difference between the two sites decreases, generally higher  $Q_H$  are measured at Tornø than Kumpula. In spring, CLM

overestimates  $Q_H$  at both sites (RMSE = 60 and 49 W m<sup>-2</sup>). SUEWS overestimates daytime  $Q_H$  at Torn  (RMSE = 64 W m<sup>-2</sup>) and underestimates the values at Kumpula (RMSE = 47 W m<sup>-2</sup>) similarly as SURFEX at both sites (RMSE = 53 and 46 W m<sup>-2</sup>). All models have difficulties in predicting the peaks in spring, which suggests problems in simulating the timing of the peak heat storage. A similar bias was also observed in the international model comparison study (Best and Grimmond, 2013).

In summer all models are able to simulate the daily pattern of  $Q_H$  well at Torn  (RMSE = 61 - 67 W m<sup>-2</sup>), but SUEWS underestimates the nocturnal  $Q_H$  by 20 W m<sup>-2</sup>. Whereas at Kumpula, SURFEX is the only model to successfully simulate the overall magnitude of the flux even though its peak lags two hours behind the observed peak (RMSE = 50 W m<sup>-2</sup>). CLM grossly overestimates particularly daytime flux (RMSE = 68 W m<sup>-2</sup>) and SUEWS underestimates the flux during the day (RMSE = 53 W m<sup>-2</sup>). In autumn, CLM systematically overestimates and SUEWS and SURFEX underestimate  $Q_H$  at Torn . At Kumpula all models perform better, but underestimate the daytime  $Q_H$  and overestimate the nocturnal values.

Although some of the differences in model performances for simulating  $Q_H$  are attributable to the use of different parameters, the turbulent flux calculation methods also vary. SURFEX and CLM use a resistance method based on the temperature difference between air and wall, roof and canyon floor (Table 2), whereas in SUEWS  $Q_H$  is calculated as a residual from the other energy balance terms. The wintertime overestimation of  $Q_H$  by CLM is largely explained by the thermal surface properties and the same properties could also affect the underestimation of  $Q_H$  by SURFEX. Other possible sources for the underestimation could be issues in simulating correct

surface temperatures, waste heat or aerodynamic resistances. SURFEX has been found to be sensitive particularly for the aerodynamic resistance between the canyon and above air (Lemonsu *et al.*, 2004), which was also supported by the current study (not shown). The resistances between the wall and canyon air and road and canyon air are an unlikely reason for the underestimation, as the differences in simulating friction velocities (further affecting aerodynamic resistances) cannot have the observed effect. As  $Q_H$  depends also on the internal building temperature ( $T_{air}$  in the case of SUEWS), too high set-point temperature limit reduces both  $Q_F$  and  $Q_H$ . Unfortunately, no independent data are available to verify the modelled  $Q_F$ . For CLM, the internal building temperature value was found to be less important than the effect of thermal properties (as discussed above). In SUEWS the differences between the observed and modelled  $Q_H$  are related to problems in simulating other energy balance components as it is calculated as a residual.

Taylor diagrams for  $Q_H$  at Tornø and Kumpula (Figures 5a,c) show that none of the studied models is able to outperform the other models by means of systematically higher correlation, lower normalized RMSE and standard deviation close to one, but rather the performance varies with season and model. At Tornø, the correlations vary between 0.5 - 0.83, normalized RMSE between 0.6 - 0.9 and normalized standard deviation between 0.35 - 1.2. Similarly at Kumpula, the correlations are between 0.6 - 0.87, normalized RMSE between 0.5 - 0.9 and normalized standard deviations between 0.4 - 1.2. Overall  $Q_H$  is simulated better at Kumpula than at Tornø, as the correlation values are higher and normalised RMSE generally lower at Kumpula. At the more densely built and populated site Tornø, problems related to thermal properties and anthropogenic heat emissions during the heating season in winter will

affect more strongly  $Q_H$  than at the more vegetated Kumpula site. This suggests that both climate and surface type have an equal effect on model performances.

#### 5.4 Latent heat flux

Large variability also occurs when simulating  $Q_E$  (Figure 6). In winter, when the vegetation is inactive, similar diurnal behaviour of observed  $Q_E$  at both sites is observed. CLM and SUEWS simulate the diurnal cycle well at both sites and RMSEs range between 11 and 22  $\text{W m}^{-2}$ . Whereas SURFEX underestimates  $Q_E$  both at Tornio and Kumpula (RMSE = 17 and 16  $\text{W m}^{-2}$ ). In spring and summer, the larger surface fraction of vegetation at Kumpula results in 20 % and 38 %, respectively, higher observed daytime  $Q_E$  than observed at Tornio. In spring, SUEWS models the diurnal behaviour well at Kumpula (RMSE = 26  $\text{W m}^{-2}$ ), but otherwise all models underestimate  $Q_E$  at both sites resulting in RMSE = 28 – 36  $\text{W m}^{-2}$ . Spring and early summer are periods for intensive leaf growth (Figure 7a, c) and therefore the seasonal changes in LAI are a possible source for uncertainty especially in the case of  $Q_E$ . Both CLM and SURFEX use fixed LAI profile for each year, whereas in SUEWS this dynamically changes with  $T_{air}$ . CLM and SURFEX were also run using LAI calculated by SUEWS, but this had only a minor effect to  $Q_E$  (not shown).

Underestimation of  $Q_E$  continues in summer with SURFEX getting the daytime peaks closest to the observations at both sites. CLM underestimates  $Q_E$  more than 50% at Kumpula with RMSE = 60  $\text{W m}^{-2}$ , while for SUEWS and SURFEX the RMSEs are 47 and 42  $\text{W m}^{-2}$ . In summer, irrigation can play an important role in urban water balance, and e.g. in SUEWS this is a default model feature. In CLM, no irrigation by



default is used, whereas in SURFEX, none of the vegetation types used in Helsinki (Table 1) have irrigation. Irrigation can be particularly important at Kumpula, where the allotment and university botanical gardens are located and intensive watering can take place. At Tornio, the small surface fraction of vegetation and minor irrigation in public parks and street trees diminish its importance. This study site dependency was also visible when SUEWS was run without irrigation resulting in increased RMSE from 47 to 58 W m<sup>-2</sup> at Kumpula and decreased RMSE from 49 to 38 W m<sup>-2</sup> at Tornio. Thus, some of the underestimation of CLM and SURFEX at least at Kumpula can be explained by the missing irrigation. For CLM, also global simulations of summer latent heat in the high-latitude regions compared to observation-based estimates (Jung *et al.*, 2011) do indicate some negative biases with regards to latent heat from vegetation, which is consistent with the results found here. A more recent version of CLM has increased latent heat at high-latitudes due to revisions in model structure and photosynthetic parameters (Bonan *et al.*, 2011).

Models systematically underestimate the nocturnal values in summer by 15 W m<sup>-2</sup>. Similar strength of nocturnal  $Q_E$  has been observed, e.g. in Melbourne (Coutts *et al.*, 2007). A possible explanation could be underestimation of nocturnal anthropogenic and storage heat emissions to the urban atmosphere. In autumn, the observed difference between the two sites again diminishes and the models again underestimate  $Q_E$ , and the RMSEs range between 21 and 29 W m<sup>-2</sup>.

Taylor-diagram indicates that again the models are able to simulate  $Q_E$  better at Kumpula than at Tornio as the correlations are, respectively, between 0.45 - 0.84 and 0.2 - 0.6 and normalized RMSE between 0.45 - 1.1 and 0.8 - 1.3 (Figure 5b,d).

Notably  $Q_E$  is more problematic for all models than  $Q^*$  or  $Q_H$  (Table 3 and 4, Figure 5), which is consistent with previous evaluations of urban land surface models in mid-latitude cities (Grimmond *et al.*, 2011). The underestimation of  $Q_E$  can be related to the description of vegetation, which is affected both by LAI and surface cover fractions. A recent study shows how a decrease in pixel resolution of surface cover fractions from 2 to 100 m can increase the mean bias of models by 50% (Nordbo *et al.*, 2015). At Torni, the underestimation can also be related to anthropogenic water emissions from traffic and respiration of people seen in the observations, as the model runs do not account for these. In cold climates, these sources of moisture are proportionally more significant given the typically dry natural specific humidity conditions. However, the surface resistances are a key driver in all models to calculate evaporation and related energy from urban surface and therefore future work in urban areas is clearly needed to investigate these as a potential source for the bias in  $Q_E$ .

### 5.5 The effect of measurement uncertainty

The normalized acceptable deviation ( $D$ ) for  $20 \text{ W m}^{-2}$  bins is plotted in Figure 8. The plotted distributions can be used to study the seemingly poor model performances when simulating e.g. diurnal cycles. For example, CLM tends to overestimate  $Q_H$  at Kumpula throughout the year, but with large observed values ( $>230 \text{ W m}^{-2}$ ) the modelled values are within the measurement uncertainties 80% of the time (Figure 8c). This suggests that the CLM simulates  $Q_H$  well within the uncertainties of the EC measurements in daytime in summer despite the evident overestimation seen in Figure 5g. Thus the overestimation of  $Q_H$  by CLM is considerable only with small  $Q_H$  corresponding wintertime and nocturnal values. Similarly, all models underestimate

$Q_E$  at both sites in daytime during summer. For SURFEX and SUEWS the modelled values are 70% and 40-65%, respectively, within the measurement uncertainties at Kumpula when the observed  $Q_E > 100 \text{ W m}^{-2}$ . For CLM, less than 20% of the modelled  $Q_E$  are within the uncertainties of observations. At Tornio, modelled  $Q_E$  is less than 35% of the times within the measurement uncertainties with observed  $Q_E > 100 \text{ W m}^{-2}$ . The underestimation can be explained by the small frequency of large observed  $Q_E$  values in the Helsinki city centre.

### 5.6 Surface energy partitioning

The complete surface energy balance varies between the models, throughout the year and relative to the monthly means for observations of  $Q^*$ ,  $Q_H$  and  $Q_E$  (Figure 9). From the latter, some indication of the uncertainty of the energy partitioning for high-latitude cities can be considered. In mid-latitude cities,  $Q^*$  is typically the main component fuelling the turbulent exchanges (Grimmond et al., 2011), but in Helsinki this is the case only for the period from April to September. During winter months, both  $Q_F$  and  $\Delta Q_s$  become more important than  $Q^*$  due to small incoming solar radiation. The increased importance of  $Q_F$  in winter months is explained partially by the decrease of  $Q^*$ , and partially by the annual variability of  $Q_F$  driven by the heating need in Helsinki.

As was already evident from the diurnal cycles (Figure 3), the differences in monthly  $Q^*$  between the different models are rather small. At Tornio, CLM gives throughout the year systematically lower  $Q_F$  ( $1 - 29 \text{ W m}^{-2}$ ) than the other two models ( $26 - 45 \text{ W m}^{-2}$  and  $17 - 46 \text{ W m}^{-2}$  for SUEWS and SURFEX, respectively). The near-zero  $Q_F$  in

summer simulated by CLM is due to lack of building cooling, energy consumption and heat emissions from traffic. Both SUEWS and SURFEX take heat emissions from traffic into account by using a constant  $Q_F$  traffic profile and constant single value of  $10 \text{ W m}^{-2}$ , respectively. Cooling of buildings is only considered in SUEWS and energy consumption only in SURFEX. In summer, the higher  $Q_F$  given by SUEWS than given by SURFEX (35% in July) can be explained by the cooling of buildings. In winter, both SURFEX and SUEWS give almost the same  $Q_F$  ( $42 \text{ W m}^{-2}$  in January). This corresponds well with the rough wintertime estimate of  $Q_F = 50 \text{ W m}^{-2}$  obtained from the EC measurements (Nordbo *et al.*, 2013). At Kumpula, the anthropogenic heat emissions are smaller than at Tornio due to the less built-up and populated surface. CLM and SUEWS give lower  $Q_F$  than SURFEX, with the monthly means ranging between  $0 - 10$ ,  $5 - 14$  and  $16 - 27 \text{ W m}^{-2}$ , respectively (Figure 9).

The annual variability of  $\Delta Q_s$  is distinct at both sites, but all models give slightly different behaviour. SUEWS gives the largest heat storage at both sites in May and July with slightly higher values at Tornio ( $34 - 37 \text{ W m}^{-2}$ ) than at Kumpula ( $22 - 24 \text{ W m}^{-2}$ ). SURFEX on the other hand gives the largest heat storage (around  $14 \text{ W m}^{-2}$ ) at Kumpula in April and May, whereas at Tornio only  $5 \text{ W m}^{-2}$  in May is modelled. With CLM, the heat storages are small and maxima of  $10$  and  $7 \text{ W m}^{-2}$  are observed at Tornio and Kumpula, respectively, in July. CLM gives the largest storage heat emissions to the atmosphere at both sites in January and December with the values reaching  $-124 \text{ W m}^{-2}$  at Tornio and  $-43 \text{ W m}^{-2}$  at Kumpula. SURFEX gives the largest storage heat emissions at the same time reaching  $-23 \text{ W m}^{-2}$  and  $-13 \text{ W m}^{-2}$  at the two sites, whereas SUEWS gives the maxima storage heat emissions already in October and November reaching  $-3$  and  $-23 \text{ W m}^{-2}$  at Tornio and Kumpula, respectively. The

large storage term in CLM in winter can be explained by the calculation of  $\Delta Q_s$  as a residual from the other components of which  $Q_H$  is largely overestimated (Figures 4 and 9).

Most of the available energy is converted to  $Q_H$  particularly at Tornio where the fraction of vegetation is smaller (Figure 9). The energy partitioning between  $Q_H$  and  $Q_E$  is visualized in Figure (7b, d), where a 28-day running mean of Bowen ratio ( $\beta = Q_H/Q_E$ ) for both sites over the entire day is plotted. The observed  $\beta$  range between 1 and 8 at Tornio and 0 and 4 at Kumpula, with higher values outside the growing season when trees are without leaves. The heating period (December to mid-March) has higher  $\beta$ . Also, the start and end of the growing season are seen as a strong decrease in early May and increase in mid-October at Tornio. At Kumpula such strong changes are not seen. A possible explanation for this difference is that in the centre of Helsinki trees are mainly lime trees so there are little phenological differences in leaf growth. At Kumpula, there is wider range of vegetation species so leaf-on is more variable. Using the changes in  $\beta$  at Tornio as a proxy for LAI, the timing for the leaf-growth and senescence are well simulated by SUEWS and correctly prescribed in SURFEX (Figure 7a,c).

During the winter, all models get the general behaviour of  $\beta$  correctly despite the problems related to the magnitude of  $Q_H$  and  $Q_E$  (Figure 7b, d). However, the underestimation of  $Q_E$  by all models in the city centre in spring and summer results in too high  $\beta$ . SUEWS and SURFEX simulate values of 5 and 8 for  $\beta$  when the observed are between 1 and 5, and CLM gives as high as  $\beta = 15$  in April. None of the models simulates the decrease in  $\beta$  due to leaf growth, whereas the timing of the senescence

in late August, early September is seen with CLM and SUEWS. At Kumpula, SUEWS and SURFEX simulate the energy partitioning well during the growing season, but CLM again has a too high  $\beta$ .

### 5.7 Modelling wintertime near-surface atmospheric stability

Stable conditions are a critical concern for cold climate cities when shallow boundary layers can lead to very poor air quality. Figure 10 summarizes the occurrence of different stability conditions near the surface calculated from observations and model outputs for winter period, when stable boundary layers are most common in Helsinki. The atmospheric stabilities are calculated according to (e.g. Seinfeld and Pandis (1997))

$$\zeta = \frac{1}{L} = \frac{\frac{g}{\bar{\theta}} \overline{w'T'}}{\frac{u_*^3}{k}}, \quad (5)$$

where  $L$  is the Obukhov length (m),  $g$  is the gravitational acceleration ( $\text{m s}^{-2}$ ),  $\bar{\theta}$  is mean potential temperature (K),  $\overline{w'T'}$  is the kinematic heat covariance ( $\text{Km s}^{-1}$ ),  $u^*$  is the friction velocity ( $\text{m s}^{-1}$ ) and  $k$  is dimensionless von Karman constant. The different stability classes are (Seinfeld and Pandis, 1997): very stable ( $\zeta > 0.1$ ), stable ( $0.0001 \leq \zeta \leq 0.1$ ), neutral ( $|\zeta| < 0.00001$ ), unstable ( $-0.01 \leq \zeta \leq -0.0001$ ) and very unstable ( $\zeta < -0.01$ ).

The observed nocturnal conditions are more unstable at Tornio (Figure 10a) than at Kumpula (Figure 10e) following the differences in  $Q_H$ . At Tornio, the occurrence of stable cases is below 15% and at Kumpula 39%. Similarly, very unstable cases are frequently measured at Tornio, reaching 38% in daytime, whereas at Kumpula very

stable is only observed less than 11% of the time. All models are able to simulate the differences in  $\zeta$  between the two sites, but differences in the performances are clear.

CLM and SUEWS underestimate the number of nocturnal stable cases at Kumpula and the occurrences reach 18 and 26 %, respectively. SURFEX on the other hand overestimates the occurrence of stable cases with the occurrence reaching 47%. At Tornio, SURFEX also overestimates the number of stable cases (maximum occurrence of 21%), SUEWS clearly overestimates their occurrence (maximum 25%) and CLM completely misses them due to the too large simulated  $Q_H$  (Figure 4a). SUEWS also simulates too many very stable cases at both sites at night-time reaching an occurrence of 7% at Kumpula. All models only slightly overestimate the number of neutral cases, except CLM at Tornio, and the occurrence of neutral cases range between 1% (CLM and SUEWS ) and 5% (SURFEX) at Kumpula when compared to the observed  $\zeta$ . SUEWS and CLM also overestimate the number of very unstable cases at both sites and SURFEX underestimates at Tornio but overestimates at Kumpula. These results show how all models are able to simulate these differences in  $\zeta$  between the two sites, but in order to improve the description of LSM in air quality models more work related particularly to wintertime  $Q_H$  is needed.

## 6. Conclusions

In this study three different urban land surface models (CLM, SUEWS and SURFEX) are compared offline with observations collected in Helsinki, Finland. Models are run for two study areas with different land uses: a highly built-up city centre (78%) and a heterogeneous suburban site (42 - 54%). The main focus is on the seasonal and

diurnal changes of the surface energy balance components due to changes in snow cover, leaf area index and anthropogenic heat emissions.

To a reasonable degree, the models are able to account for the differences between the study sites as well as the intra-annual and diurnal variability of the energy balance components. None of the models outperforms others, but rather one model is closer to the observations in a component for a particular season. Previously, the international urban model comparison study observed systematically lower model performances in summer in a mid-latitude city (Best and Grimmond, 2013), whereas we have found that in Helsinki lowered model performances link also to seasonality in snow and solar radiation.

Most of the uncertainties in simulating  $Q^*$  relate to the timing of the snowmelt in spring (inferred by snow depth), when snow covered ground fraction plays an important role. Otherwise, the models simulate the formation and development of the snow pack well despite the differences in the snow parameterizations and the effect on  $Q^*$  seem to be small. Larger variability in the model performances is seen for  $Q_H$  and  $Q_E$  than for  $Q^*$  in accordance with previous model comparison studies. The largest problem in simulating  $Q_H$  relate to  $Q_F$  and thermal surface parameters affecting particularly  $\Delta Q_s$  in winter when the importance of  $Q^*$  is small due to small incoming solar radiation.  $Q_H$  bias also results in problems in simulating the near-surface stabilities particularly at the dense city centre where the built surface fraction is high and anthropogenic heat emissions greater. However, in most cases all three models catch the diurnal variability of stabilities by even creating too many very stable conditions. With the exception of CLM, which only generates unstable surface fluxes



at Tornio. Models have the most difficulties in simulating  $Q_E$ . Missing irrigation in CLM and SURFEX can explain some of the decreased model performances at the more vegetated Kumpula site, whereas at Tornio the problems are more likely related to the description of vegetation.

Although CLM and SURFEX parameterize most of the surface energy balance components similarly (when compared to SUEWS) it is surprising to see such large differences between the two models. Thus, partly these differences arise from the surface radiative and thermal properties obtained from the ECOCLIMAP and Jackson *et al.* (2010) databases. In particular, the parameters in the Jackson *et al.* (2010) database should be revisited for cold climate cities. Also for the natural surface fraction, ISBA and CLM have large differences including the processes and parameters attributed to plant types. Despite SUEWS parameterizing many processes differently to CLM and SURFEX, it gave more similar results to SURFEX than CLM.

Unfortunately most of the uncertainties relate to processes that are difficult to measure directly; i.e. melting of snow, heat storage and anthropogenic heat fluxes. Thus, their description in urban land-surface models needs to be improved in order to simulate the urban surface energy balance correctly particularly in cold climate cities. More measurements of the surface fluxes are needed in order to cover wide range of different cities varying by their cultural heritage, heating systems and architecture. Furthermore, the evapotranspiration in urban areas is poorly parameterized and therefore special attention should be given to its parameterization in urban areas. From these results it is clear that not only do different parameters need to be considered in cities in cold climates and at high-latitudes compared to those in mid-

latitude cities but also that land use (urban, suburban) can have a large effect on the model performances.

### Acknowledgements

We thank the Academy of Finland (Project numbers 138328, 1127756, ICOS 271878, ICOS-Finland 281255, ICOS-Finland 263149 and ICOS-ERIC 281250), the Academy of Finland Centre of Excellence (project no 1118615) and EU FP7 project emBRACE (283201) for their support. The work of M. Demuzere is supported by the Flemish regional government through a FWO post-doctoral position. We also express our thanks to Erkki Siivola and Sami Haapanala for instrument maintenance and Sokos Hotel Torni for providing the measurement platform in the centre of Helsinki.

### Supporting Information

Figure S1. Modelled and observed diurnal cycles of outgoing shortwave radiation ( $K \uparrow$ ).

Figure S2. Modelled and observed diurnal cycles for the outgoing longwave radiation ( $L \uparrow$ ).

### References

- Aubinet M, Vesala T, Papale D. 2012. *Eddy Covariance: A Practical Guide to Measurement and Data Analysis*. Springer Atmospheric Sciences, Netherlands.
- Best MJ, Grimmond CSB. 2013. Analysis of the Seasonal Cycle Within the First International Urban Land-Surface Model Comparison. *Bound.-Lay. Meteorol.* 146: 421 - 446. DOI: 10.1007/s10546-012-9769-7

- Bonan GB, Lawrence PJ, Oleson KW, Levis S, Jung M, Reichstein M, Lawrence DM, Swenson SC. 2011. Improving canopy processes in the Community Land Model version 4 (CLM4) using global flux fields empirically inferred from FLUXNET data. *Biogeosciences* 116. DOI: 10.1029/2010jg001593
- Boone A, Masson V, Meyers T, Noilhan J. 2000. The influence of the inclusion of soil freezing on simulations by a soil-vegetation-atmosphere transfer scheme. *J. Appl. Meteorol.* 39: 1544 - 1569. DOI: 10.1175/1520-0450(2000)039<1544:tiotio>2.0.co;2
- Coutts AM, Beringer J, Tapper NJ. 2007. Impact of increasing urban density on local climate: Spatial and temporal variations in the surface energy balance in Melbourne, Australia. 46: 477 - 493. DOI: 10.1175/jam2462.1
- Demuzere M, Oleson K, Coutts AM, Pigeon G, van Lipzig NPM. 2013. Simulating the surface energy balance over two contrasting urban environments using the Community Land Model Urban. *Int. J. Climatol.* 33: 3182 - 3205. DOI: 10.1002/joc.3656
- Douville H, Royer JF, Mahfouf JF. 1995. A new snow parameterization for the Meteo-France climate model, Part 1 - Validation in stand-alone experiments. *Clim. Dynam.* 12: 21 - 35. DOI: 10.1007/bf00208760
- Faroux S, Tchuente ATK, Roujean JL, Masson V, Martin E, Le Moigne P. 2013. ECOCLIMAP-II/Europe: a twofold database of ecosystems and surface parameters at 1 km resolution based on satellite information for use in land surface, meteorological and climate models. *Geosci. Model Dev.* 6: 563 - 582. DOI: 10.5194/gmd-6-563-2013
- Grimmond CSB, Blackett M, Best MJ, Baik JJ, Belcher SE, Beringer J, Bohnenstengel SI, Calmet I, Chen F, Coutts A, Dandou A, Fortuniak K,

- Gouvea ML, Hamdi R, Hendry M, Kanda M, Kawai T, Kawamoto Y, Kondo H, Krayenhoff ES, Lee SH, Loridan T, Martilli A, Masson V, Miao S, Oleson K, Ooka R, Pigeon G, Porson A, Ryu YH, Salamanca F, Steeneveld GJ, Tombrou M, Voogt JA, Young DT, Zhang N. 2011. Initial results from Phase 2 of the international urban energy balance model comparison. *Int. J. Clim.* 31: 244 - 272. DOI: 10.1002/joc.2227
- Grimmond CSB, Blackett M, Best MJ, Barlow J, Baik JJ, Belcher SE, Bohnenstengel SI, Calmet I, Chen F, Dandou A, Fortuniak K, Gouvea ML, Hamdi R, Hendry M, Kawai T, Kawamoto Y, Kondo H, Krayenhoff ES, Lee SH, Loridan T, Martilli A, Masson V, Miao S, Oleson K, Pigeon G, Porson A, Ryu YH, Salamanca F, Shashua-Bar L, Steeneveld GJ, Tombrou M, Voogt J, Young D, Zhang N. 2010. The International Urban Energy Balance Models Comparison Project: First Results from Phase 1. *J. Appl. Meteorol. Clim.* 49: 1268 - 1292. DOI: 10.1175/2010jamec2354.1
- Grimmond CSB, Cleugh HA, Oke TR. 1991. An objective urban heat-storage model and its comparison with other schemes. *Atmos. Environ.* 25: 311 - 326. DOI: 10.1016/0957-1272(91)90003-w
- Grimmond CSB, Oke T. 1991. An evaporation-interception model for urban areas. *Water Resour. Res.* 27: 1739 - 1755.
- Hollinger DY, Richardson AD. 2005. Uncertainty in eddy covariance measurements and its application to physiological models. *Tree Physiol.* 25: 873 - 885.
- Holtslag AAM, Svensson G, Baas P, Basu S, Beare B, Beljaars ACM, Bosveld FC, Cuxart J, Lindvall J, Steeneveld GJ, Tjernstrom M, Van de Wiel BJH. 2013. Stable atmospheric boundary layers and diurnal cycles-Challenges for

Weather and Climate Models. *B. Am. Meteorol. Soc.* 94: 1691 - 1706. DOI: 10.1175/bams-d-11-00187.1

Hurrell JW, Holland MM, Gent PR, Ghan S, Kay JE, Kushner PJ, Lamarque JF, Large WG, Lawrence D, Lindsay K, Lipscomb WH, Long MC, Mahowald N, Marsh DR, Neale RB, Rasch P, Vavrus S, Vertenstein M, Bader D, Collins WD, Hack JJ, Kiehl J, Marshall S. 2013. The Community Earth System Model A Framework for Collaborative Research. *B. Am. Meteorol. Soc.* 94: 1339 - 1360. DOI: 10.1175/bams-d-12-00121.1

IPCC. 2013. '*Climate Change 2013: The Physical Science Basis. Contribution of Working group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*,' C. U. Press: Cambridge, UK, and New York, NY, USA

Jackson TL, Feddema JJ, Oleson KW, Bonan GB, Bauer JT. 2010. Parameterization of Urban Characteristics for Global Climate Modeling. *Ann. Assoc. Am. Geogr.* 100: 848 - 865.

Järvi L, Grimmond CSB, Christen A. 2011. The Surface Urban Energy and Water Balance Scheme (SUEWS): Evaluation in Los Angeles and Vancouver. *J. Hydrol.* 411: 219 - 237. DOI: 10.1016/j.jhydrol.2011.10.001

Järvi L, Grimmond CSB, Taka M, Nordbo A, Setälä H, Strachan IB. 2014. Development of the Surface Urban Energy and Water Balance Scheme (SUEWS) for cold climate cities. *Geosci. Model Dev.* 7: 1691 - 1711. DOI: 10.5194/gmd-7-1691-2014

Järvi L, Nordbo A, Junninen H, Riikonen A, Moilanen J, Nikinmaa E, Vesala T. 2012. Seasonal and annual variation of carbon dioxide surface fluxes in Helsinki, Finland, in 2006–2010. *Atmos. Chem. Phys.* 12: 8475 - 8489. DOI: 10.5194/acp-12-8475-2012

- Jung M, Reichstein M, Margolis HA, Cescatti A, Richardson AD, Arain MA, Arneth A, Bernhofer C, Bonal D, Chen J, Gianelle D, Gobron N, Kiely G, Kutsch W, Lasslop G, Law BE, Lindroth A, Merbold L, Montagnani L, Moors EJ, Papale D, Sottocornola M, Vaccari F, Williams C. 2011. Global patterns of land-atmosphere fluxes of carbon dioxide, latent heat, and sensible heat derived from eddy covariance, satellite, and meteorological observations. *J. Geophys. Res.* 116. DOI: 10.1029/2010jg001566
- Lawrence DM, Oleson KW, Flanner MG, Thornton PE, Swenson SC, Lawrence PJ, Zeng X, Yang Z-L, Levis S, Sakaguchi K, Bonan GB, Slater AG. 2011. Parameterization Improvements and Functional and Structural Advances in Version 4 of the Community Land Model. *J. Adv. Mod. Earth Sys.* 3: 1 - 27. DOI:10.1029/2011ms000045
- Lemonsu A, Bélair S, Mailhot J, Leroyer S. 2010. Evaluation of the Town Energy Balance Model in Cold and Snowy Conditions during the Montreal Urban Snow Experiment 2005. *J. Appl. Meteorol. Clim.* 49: 346 - 362. DOI: 10.1175/2009jamc2131.1
- Lemonsu A, Grimmond CSB, Masson V. 2004. Modeling the surface energy balance of the core of an old Mediterranean city: Marseille. 43: 312-327. DOI: 10.1175/1520-0450(2004)043<0312:mtsebo>2.0.co;2
- Lemonsu A, Masson V, Shashua-Bar L, Erell E, Pearlmutter D. 2012. Inclusion of vegetation in the Town Energy Balance model for modelling urban green areas. *Geosci. Model Dev.* 5: 1377 - 1393. DOI: 10.5194/gmd-5-1377-2012
- Leroyer S, Mailhot J, Bélair S, Lemonsu A, Strachan IB. 2010. Modeling the Surface Energy Budget during the Thawing Period of the 2006 Montreal Urban Snow

Experiment. *J. Appl. Meteorol. Clim.* 49: 68 - 84. DOI:

10.1175/2009jamc2153.1

Loridan T, Grimmond CSB, Offerle BD, Young DT, Smith TEL, Järvi L, Lindberg F.

2011. Local-Scale Urban Meteorological Parameterization Scheme (LUMPS):

Longwave Radiation Parameterization and Seasonality-Related

Developments. *J. Appl. Meteorol. Clim.* 50: 185 - 202. DOI:

10.1175/2010jamc2474.1

Masson V. 2000. A physically-based scheme for the urban energy budget in

atmospheric models. *Bound.-Lay. Meteorol.* 94: 357 - 397. DOI:

10.1023/a:1002463829265

Masson V, Champeaux JL, Chauvin F, Meriguet C, Lacaze R. 2003. A global

database of land surface parameters at 1-km resolution in meteorological and

climate models. *J. Climate* 16: 1261 - 1282. DOI: 10.1175/1520-0442-

16.9.1261

Masson V, Le Moigne P, Martin E, Faroux S, Alias A, Alkama R, Belamari S, Barbu

A, Boone A, Bouyssel F, Brousseau P, Brun E, Calvet JC, Carrer D,

Decharme B, Delire C, Donier S, Essaouini K, Gibelin AL, Giordani H,

Habets F, Jidane M, Kerdraon G, Kourzeneva E, Lafaysse M, Lafont S,

Lebeaupin Brossier C, Lemonsu A, Mahfouf JF, Marguinaud P, Mokhtari M,

Morin S, Pigeon G, Salgado R, Seity Y, Taillefer F, Tanguy G, Tulet P,

Vincendon B, Vionnet V, Voldoire A. 2013. The SURFEXv7.2 land and

ocean surface platform for coupled or offline simulation of earth surface

variables and fluxes. *Geosci. Model Dev.* 6: 929 - 960. DOI: 10.5194/gmd-6-

929-2013

- McMichael AJ, Woodruff RE, Hales S. 2006. Climate change and human health: present and future risks. *Lancet* 367: 856-869. DOI:10.1016/S0140-6736(06)68079-3
- Noilhan J, Planton S. 1989. A simple parameterization of land surface processes for meteorological models. *Mon. Weather Rev.* 117: 536 - 549. DOI: 10.1175/1520-0493(1989)117<0536:aspols>2.0.co;2
- Nordbo A, Järvi L, Haapanala S, Moilanen J, Vesala T. 2013. Intra-City Variation in Urban Morphology and Turbulence Structure in Helsinki, Finland. *Bound.-Lay. Meteorol.* 146: 469 - 496. DOI: 10.1007/s10546-012-9773-y
- Nordbo A, Järvi L, Vesala T. 2012. Revised eddy covariance flux calculation methodologies – effect on urban energy balance. *Tellus B* 64: 18184. DOI: 10.3402/tellusb.v64i0.18184
- Nordbo A, Karsisto P, Matikainen L, Wood CR, Järvi L. 2015. Urban surface cover determined with airborne lidar – implications for surface energy balance modelling. *In review in Urban Climate* 13: 52-72. DOI: 10.1016/j.uclim.2015.05.004
- Offerle B, Grimmond CSB, Oke TR. 2003. Parameterization of net all-wave radiation for urban areas. *J. Appl. Meteorol.* 42: 1157 - 1173. DOI: 10.1175/1520-0450(2003)042<1157:ponarf>2.0.co;2
- Oke TR. 1987. City form and natural processes. *Environ. Behav.* 19: 753 - 755. DOI: 10.1177/0013916587196012
- Oleson KW, Bonan GB, Feddema J, Vertenstein M. 2008a. An urban parameterization for a global climate model. Part II: Sensitivity to input parameters and the simulated urban heat island in offline Simulations. *J. Appl. Meteorol. Climatol.* 47: 1061 - 1076. DOI: 10.1175/2007jamc1598.1



- Oleson KW, Bonan GB, Feddema J, Vertenstein M, Grimmond CSB. 2008b. An urban parameterization for a global climate model. Part I: Formulation and evaluation for two cities. *J. Appl. Meteorol. Climatol.* 47: 1038 - 1060. DOI: 10.1175/2007jamc1597.1
- Oleson KW, Bonan GB, Feddema J, Vertenstein M, Kluzek E. 2010a. 'Technical Description of an Urban Parameterization for the Community Land Model (CLMU),' Technical Report NCAR/TN-480+STR. NCAR: Boulder, CO, USA.
- Oleson KW, Lawrence DM, Bonan GB, Flanner MG, Kluzek E, Lawrence PJ, Levis S, Swenson SC, Thornton PE. 2010b. 'Technical description of version 4.0 of the Community Land Model (CLM),' Technical Note NCAR/TN-478+STR. NCAR: Boulder, CO, USA.
- Seinfeld JH, Pandis PN. 1997. *Atmospheric Chemistry and Physics: From air pollution to climate change*. John Wiley & Sons, USA.
- Semádeni-Davies A, Bengtsson. 1998. Snowmelt sensitivity to radiation in the urban environment. *Hydrolog. Sci. J.* 43: 67-89. DOI: 10.1080/02626669809492103
- Shaffer SR, Chow WTL, Georgescu M, Hyde P, Jenerette GD, Mahalov A, Moustaoi M, Ruddell BL. 2015. Multiscale Modeling and Evaluation of Urban Surface Energy Balance in the Phoenix Metropolitan Area. *J. Appl. Meteorol. Climatol.* 54: 322-338. 10.1175/jamc-d-14-0051.1
- Sivak M. 2013. Air conditioning versus heating: climate control is more energy demanding in Minneapolis than in Miami. *Environ. Res. Lett* 8: 1 - 4. DOI: 10.1088/1748-9326/8/1/014050
- Stewart ID, Oke TR. 2012. Local Climate Zones for Urban Temperature Studies. *B. Am. Meteorol. Soc.* 93: 1879 - 1900. DOI: 10.1175/bams-d-11-00019.1

Taylor KE. 2001. Summarizing multiple aspects of model performance in a single diagram. *J. Geophys. Res.* 106: 7183. DOI: 10.1029/2000jd900719

Tilastokeskus 2014. Population by area in Finland.

[http://www.stat.fi/tup/suoluk/suoluk\\_vaesto\\_en.html](http://www.stat.fi/tup/suoluk/suoluk_vaesto_en.html), 21.3.2014.

Vesala T, Järvi L, Launiainen S, Sogachev A, Rannik Ü, Mammarella I, Siivola E, Keronen P, Rinne J, Riikonen A, Nikinmaa E. 2008. Surface–atmosphere interactions over complex urban terrain in Helsinki, Finland. *Tellus B* 60: 188 - 199. DOI: 10.1111/j.1600-0889.2007.00312.x

World Bank 2014. Urban development. <http://data.worldbank.org/topic/urban-development>, 11.6.2015.

Table 1. Information about the study sites (see Figure 1a) and parameters used in the model runs. Where model specific values are required these are indicated. Land cover fractions for Kumpula were calculated based on Nordbo *et al.* (2015) and for Tornio from Nordbo *et al.* (2013).

	Tornio	Kumpula		
Latitude (WGS84)	60°12'10.14'' N	60°12'10.14'' N		
Longitude (WGS84)	24°57'40.06'' E	24°57'40.06'' E		
Initial date and time of the run	1 July 2011 00:00 LT	1 July 2011 00:00 LT		
Time zone	2 UTC	2 UTC		
Measurement/modelling height (m)	60	31		
Base elevation (m)	15.2	29		
Local climate zone (LCZ)	2	6		
		Ku1	Ku2	Ku3
Study area (m <sup>2</sup> )	1 960 000	447000	782000	782000
Number of capita (# m <sup>-2</sup> ) <sub>1</sub>	0.0081	0.0031	0.0037	0.0044
Fraction of built surface	0.78	0.42	0.54	0.46
Fraction of paved surface	0.40	0.27	0.39	0.32
Fraction of buildings	0.37	0.15	0.15	0.14
Fraction of vegetation	0.22	0.58	0.46	0.54
Fraction of coniferous trees/shrubs <sub>1,2</sub>	0	0.01	0	0.01
Fraction of deciduous trees/shrubs <sub>1,2</sub>	0.15	0.38	0.2	0.29
Fraction of grass surface <sub>1,2</sub>	0.07	0.20	0.26	0.24
Fraction of boreal broadleaf deciduous trees <sub>3</sub>	0.12	0.34	0.20	0.27
Fraction of boreal needle leaf evergreen tree <sub>3</sub>	0.03	0.04	0	0.02
Fraction of C3 grass surface <sub>3</sub>	0.10	0.20	0.26	0.25
Fraction of water <sub>4</sub>	0.01	0	0.01	0
Mean building height (m)	17.9	10.4	11.5	12.6
Mean tree height (m)	8.3	9.8	8.7	8.4
Starting day for irrigation <sub>1</sub>	152	152	152	152
Ending day for irrigation <sub>1</sub>	243	243	243	243
Fraction of automatic irrigation <sub>1</sub>	0.3	0.2	0.2	0.4
Previous day $T_{air}$ (°C) <sub>1,2</sub>	21	21	21	21
Days since rain <sub>1,2</sub>	4	4	4	4
Initial $T_{soil}$ 50-230 mm (°C) <sub>5</sub>	12.7	12.7	12.7	12.7
Initial $T_{soil}$ T 230-600 mm (°C) <sub>5</sub>	9.4	9.4	9.4	9.4
Init. soil water storage 0-230 mm (mm) <sub>5</sub>	54	54	54	54
Init. soil water storage 0-370 mm (mm) <sub>5</sub>	125	125	125	125
Init. surface state (mm) <sub>1</sub>	0	0	0	0

<sup>1</sup>SUEWS

<sup>2</sup>SURFEX

<sup>3</sup>CLM

<sup>4</sup>Water surface only considered in SUEWS. Water fraction replaces the respective vegetation fraction.

<sup>5</sup>Obtained from soil observations conducted at a park in central Helsinki.

Table 2. Model and equations most relevant for the current study.  $\alpha_{\text{snow}}$  = snow albedo,  $\epsilon_{\text{snow}}$  = snow emissivity,  $s_d$  = snow depth,  $f_{s,i}$  = snow fraction of  $i$ th surface,  $z_{\text{top}}$  and  $z_{\text{bot}}$  = Plant Functional Type (PFT)-specific canopy top and bottom heights,  $z_c$  = critical snow depth when short vegetation is assumed to be completely buried by snow,  $W_s$  = snow water equivalent,  $W_{s,\text{max}}$  = snow water equivalent when the surface is assumed to be completely buried by snow,  $z_0$  = roughness length, NARP = the Net all-wave Radiation Parameterization scheme and  $M$  = amount of snowmelt. See text for description of other terms.

	CLM	SUEWS	SURFEX
Version	4.0	2014b	7.2
$T_{\text{min, building}} (^{\circ}\text{C})$	19.0 <sub>1</sub>	18.2 <sub>2</sub>	19.0 <sub>1</sub>
$\alpha_{\text{snow}}$	0.66/0.56 <sub>3</sub>	0.18 – 0.85 <sub>4</sub>	0.30 – 0.85 on roofs 0.15 – 0.85 on roads
$\epsilon_{\text{snow}}$	0.97	0.99	0.98
$f_{s,\text{pav}}$	$\frac{s_d}{0.05} \leq 1$	$\left(\frac{W_s}{W_{s,\text{max}}}\right)^2$	$\frac{W_s}{(W_s + W_{s,\text{max}})}$
$f_{s,\text{roof}}$	$\frac{s_d}{0.05} \leq 1$	$0.5 \cdot \left(\frac{W_s}{W_{s,\text{max}}}\right), \frac{W_s}{W_{s,\text{max}}} < 0.9$	$\frac{W_s}{(W_s + W_{s,\text{max}})}$
$f_{s,\text{veg}}$	$\frac{s_d - z_{\text{bot}}}{z_{\text{top}} - z_{\text{bot}}}$ for trees/shrubs $\frac{\min(s_d, z_c)}{z_c}$ for grass	$1 - \left(\frac{1}{\pi} \arccos\left(2 \frac{W_s}{W_{s,\text{max}}} - 1\right)\right)^{1.7}$	$\frac{s_d}{(s_d + 5000z_0)}$
$Q^* (\text{W m}^{-2})$	$\alpha + T_{\text{surf}}$	NARP ( $\alpha + T_{\text{surf}}$ )	$\alpha + T_{\text{surf}}$
$Q_H (\text{W m}^{-2})$	Resistance method using difference between $T_{\text{air}}$ and $T_{\text{surf}}$	Residual of the energy balance	Resistance method using difference between $T_{\text{air}}$ and $T_{\text{surf}}$ + building and traffic $Q_F$
$Q_E (\text{W m}^{-2})$	Resistance method using difference between $q_{\text{air}}$ and $q_{\text{surf}}$	Penman-Monteith equation	Resistance method using difference between $q_{\text{air}}$ and $q_{\text{surf}}$
$Q_F (\text{W m}^{-2})$	Building heating	Building heating and cooling + traffic	Building heating+ industrial activities +

			traffic
$\Delta Q_s (\text{W m}^{-2})$	Residual of the energy balance	OHM	Residual of the energy balance, driven by heat conduction through surfaces
$M (\text{mm})$	Energy balance of snow (up to 5 layers)	Degree day method based on $Q^*$ and $T_{air}$	Energy balance of snow (1 layer)

<sup>1</sup>Minimum building temperature

<sup>2</sup>Outdoor temperature

<sup>3</sup>Visible and near-infrared waveband

<sup>4</sup>Snow aging accounted for

Table 3. Model evaluation statistics for CLM, SUEWS and SURFEX, based on 12 months of observations of a) net all-wave radiation ( $Q^*$ ), b) outgoing shortwave ( $K\uparrow$ ) and c) longwave ( $L\uparrow$ ) radiations for different seasons at Tornio and Kumpula.  $r$  = Pearson correlation coefficient, RMSE = root mean square error ( $\text{W m}^{-2}$ ), MBE = mean bias error ( $\text{W m}^{-2}$ ), Mean = mean of observation ( $\text{W m}^{-2}$ ) and  $N$  = number of samples.

a) $Q^*$	season	$r$	RMSE	Tornio			$r$	RMSE	Kumpula		
				MBE	Mean	$N$			MBE	Mean	$N$
CLM	winter	0.99	16.4	-11.8	-7.1	2245	0.98	6.8	1.1	-10.4	2302
	spring	1.00	30.2	-17.2	83.3	1454	0.99	16.3	2.3	71.2	1464
	summer	1.00	34.1	-10.0	104.9	3427	1.00	15.1	1.8	98.4	3431
	autumn	0.99	10.8	-7.0	-13.4	1489	1.00	5.6	0.5	-11.7	1498
SUEWS	winter	0.99	8.9	2.1	-7.1	2245	0.98	14.1	3.2	-10.4	2302
	spring	0.99	25.1	-5.8	83.3	1454	0.99	20.8	3.7	71.2	1464
	summer	1.00	19.8	-2.9	104.9	3427	1.00	8.8	-2.9	98.4	3431
	autumn	0.99	5.5	1.4	-13.4	1489	0.99	4.7	-1.0	-11.7	1498
SURFEX	winter	0.97	21.8	-2.1	-7.1	2245	0.98	10.2	0.5	-10.4	2302
	spring	0.98	44.2	-13.2	83.3	1454	0.99	24.1	-4.1	71.2	1464
	summer	1.00	27.2	-4.0	104.9	3427	1.00	16.1	-2.0	98.4	3431
	autumn	0.99	8.9	5.0	-13.4	1489	0.99	7.2	4.1	-11.7	1498
b) $K\uparrow$											
CLM	winter	0.91	10.2	2.5	6.2	2247	0.99	6.4	-2.3	15.1	2303
	spring	0.94	25.2	14.0	17.3	1454	0.94	18.3	-5.2	30.5	1464
	summer	0.92	29.2	15.8	20.4	3476	0.99	4.4	-1.3	29.0	3431
	autumn	0.85	6.3	2.3	2.1	1527	0.96	2.5	-0.2	3.8	1498
SUEWS	winter	0.90	8.6	1.1	6.2	2247	0.96	15.0	-5.4	15.1	2303
	spring	0.82	24.3	7.5	17.3	1454	0.87	24.3	-7.2	30.5	1464
	summer	0.92	13.6	4.1	20.4	3476	0.99	4.8	-1.8	29.0	3431
	autumn	0.81	4.7	1.0	2.1	1527	0.92	3.7	-0.3	3.8	1498
SURFEX	winter	0.91	16.8	5.07	6.2	2247	0.99	13.3	-5.58	15.1	2303
	spring	0.80	31.8	11.3	17.3	1454	0.95	14.5	-3.44	30.5	1464
	summer	0.85	17.6	2.1	20.4	3476	0.97	10.2	-2.45	29.0	3431
	autumn	0.81	4.9	0.9	2.09	1527	0.93	3.5	-0.46	3.8	1498
c) $L\uparrow$											
CLM	winter	0.94	11.9	9.4	295.9	2289	0.97	5.6	1.3	290.5	2302
	spring	0.93	9.6	3.3	343.9	1464	0.95	10.9	2.9	342.1	1464
	summer	0.91	13.3	-5.9	396.4	3480	0.93	15.5	-0.4	394.9	3431
	autumn	0.94	8.1	4.4	335.7	1536	0.97	5.0	-0.3	331.5	1498
SUEWS	winter	0.99	4.9	-3.2	295.9	2289	0.96	7.1	2.2	290.5	2302
	spring	0.98	5.91	-1.7	343.9	1464	0.97	8.0	3.5	342.1	1464
	summer	0.97	7.31	-1.2	396.4	3480	0.97	9.5	4.7	394.9	3431
	autumn	0.99	3.66	-2.5	335.7	1536	0.98	4.3	1.3	331.5	1498
SURFEX	winter	0.97	7.0	-4.8	295.9	2289	0.99	4.4	0.7	290.5	2302
	spring	0.95	11.2	-5.5	343.9	1464	0.99	5.3	-1.3	342.1	1464
	summer	0.95	11.8	-0.4	396.4	3480	0.98	9.4	2.0	394.9	3431
	autumn	0.98	7.4	-5.5	335.7	1536	0.99	3.9	-2.4	331.5	1498

Table 4. Same as Table 3, but for a) sensible ( $Q_H$ ) and b) latent heat ( $Q_E$ ) fluxes.

a) $Q_H$	season	$r$	RMSE	Torni			$r$	RMSE	Kumpula		
				MBE	Mean	$N$			MBE	Mean	$N$
CLM	winter	0.57	73.2	58.0	58.4	855	0.68	26.6	7.7	24.1	846
	spring	0.81	75.8	54.6	82.8	906	0.87	62.6	30.2	71.0	894
	summer	0.81	60.7	2.4	90.5	2156	0.87	67.9	31.9	51.5	2346
	autumn	0.54	41.0	27.2	44.9	822	0.76	26.9	13.8	-2.2	628
SUEWS	winter	0.62	46.4	-8.6	58.4	855	0.64	31.8	7.3	24.1	846
	spring	0.82	63.6	7.1	82.8	906	0.82	59.8	-8.4	71.0	894
	summer	0.82	67.1	3.2	90.5	2156	0.82	52.9	-3.1	51.5	2346
	autumn	0.58	38.3	-23.0	44.9	822	0.68	27.1	7.5	-2.2	628
SURFEX	winter	0.53	59.8	-39.5	58.4	855	0.63	32.6	-17.5	24.1	846
	spring	0.73	67.4	-11.6	82.8	906	0.83	59.4	-16.0	71.0	894
	summer	0.81	65.7	8.2	90.5	2156	0.86	49.6	12.7	51.5	2346
	autumn	0.57	47.5	-38.1	44.9	822	0.68	26.5	-1.6	-2.2	628
c) $Q_E$											
CLM	winter	0.20	21.8	4.8	14.4	820	0.64	11.4	-0.4	18.3	719
	spring	0.29	33.9	-9.9	26.3	676	0.73	35.8	-24.7	44.2	811
	summer	0.34	49.7	-20.2	46.1	1938	0.76	60.3	-40.1	80.0	2121
	autumn	0.21	24.7	-2.2	21.3	785	0.50	27.4	-14.7	30.1	540
SUEWS	winter	0.50	15.3	1.2	14.4	820	0.64	15.6	1.2	18.3	719
	spring	0.31	32.6	-12.8	26.3	676	0.76	25.7	-3.7	44.2	811
	summer	0.44	48.7	-23.9	46.1	1938	0.79	46.7	-23.8	80.0	2121
	autumn	0.24	24.1	-9.6	21.3	785	0.46	27.2	-12.0	30.1	540
SURFEX	winter	0.59	16.5	-9.3	14.4	820	0.73	15.9	-12.1	18.3	719
	spring	0.43	32.0	-13.9	26.3	676	0.79	28.3	-16.0	44.2	811
	summer	0.48	45.7	-17.9	46.1	1938	0.83	41.6	-18.8	80.0	2121
	autumn	0.41	21.2	-10.3	21.3	785	0.53	29.1	-18.3	30.1	540

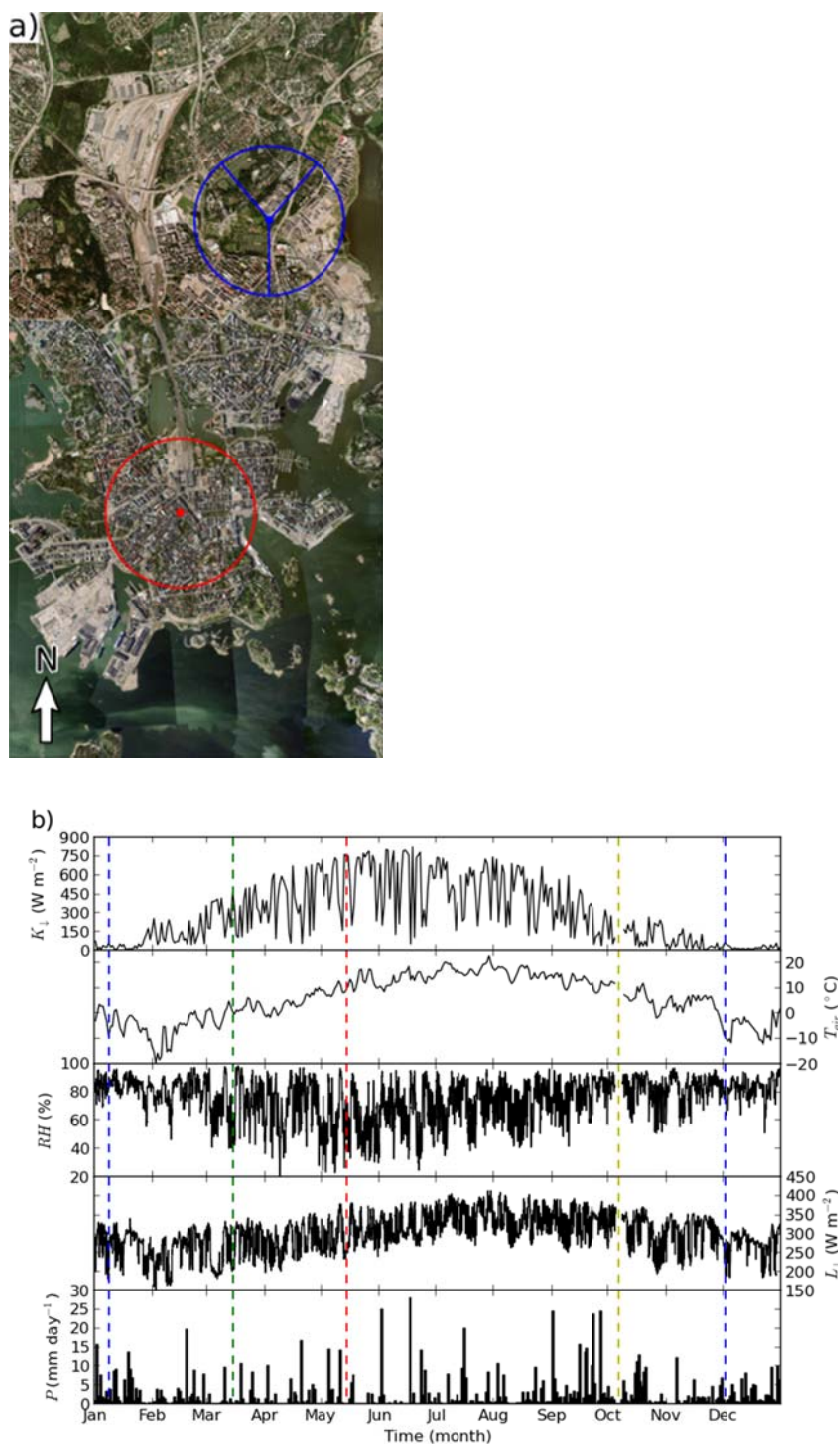


Figure 1. a) Aerial photo of Helsinki city centre (©2011 Kaupunkimittausosasto, Helsinki, Finland). Kumpula measurement site is marked with a blue dot and Hotel Torni with a red dot. Circles denote the modelled areas (radius 1 km, Table 1) and the lines at Kumpula mark the different surface cover areas (Ku1 in north, Ku2 in east



and Ku3 in west). b) Time series of selected meteorological variables observed in Kumpula in 2012. From top to bottom: daytime mean incoming shortwave radiation (average over hours 1000–1400 local time (UTC+2),  $K_{\downarrow}$ ), daily mean air temperature ( $T_{air}$ ), relative humidity (RH) and incoming longwave radiation ( $L_{\downarrow}$ ) plus daily sum of precipitation. The blue, green, red and yellow indicate the starting days for thermal (changes with year) winter, spring, summer and autumn, respectively.

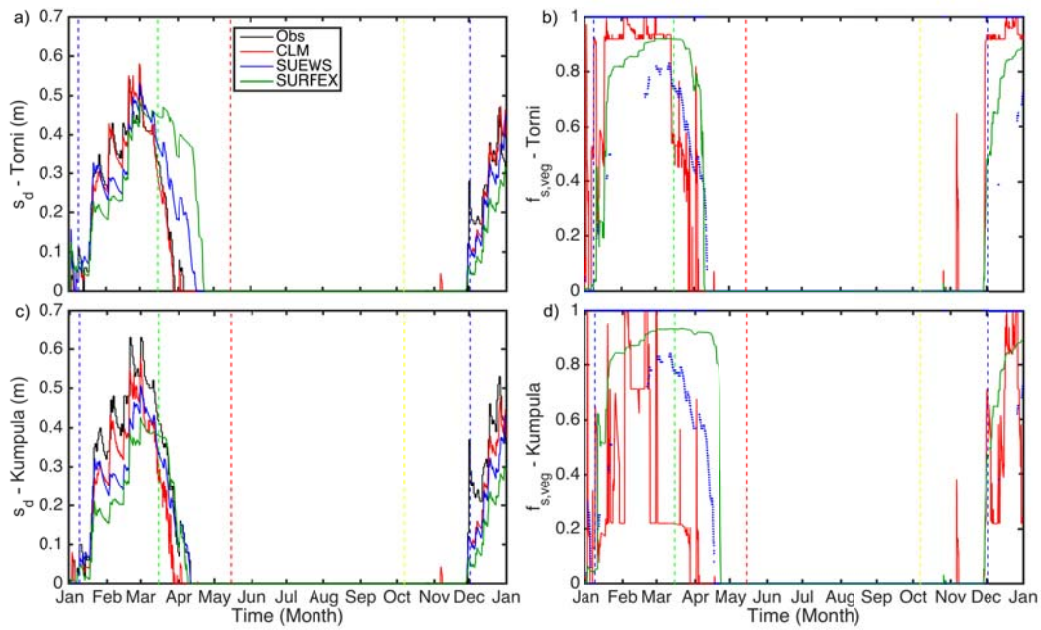


Figure 2. The observed and simulated snow depths ( $s_d$ ) at a) Tornii and c) Kumpula vegetation sector, and simulated surface fractions of snow ( $f_{s,veg}$ , Table 1) at b) Tornii and d) Kumpula vegetation sector. Snow observations are made on open areas so only  $s_d$  simulated at the vegetated tile of SURFEX and CLM, and at grass surface of SUEWS are shown. The dashed blue, green, red and yellow lines indicate the starting days for thermal (changes with year) winter, spring, summer and autumn, respectively.

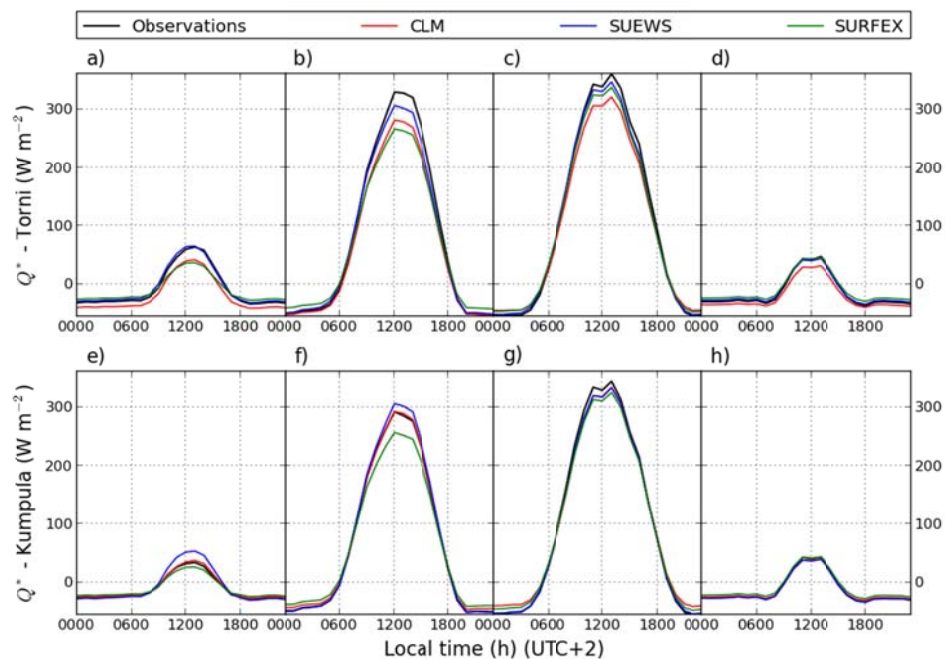


Figure 3. Modelled and observed diurnal cycles of  $Q^*$ . Subplots a-d (upper row) are for Torni in different seasons (winter, spring, summer and autumn, respectively), and subplots e-h (lower row) for Kumpula vegetation sector (Ku3) which best describes the surface cover within the source area of radiation sensor.

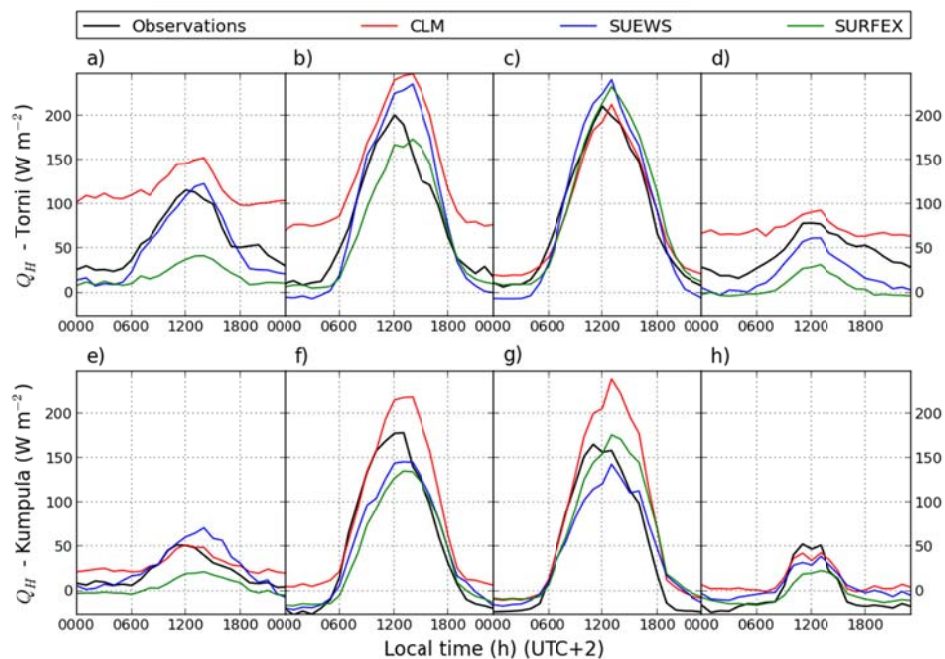
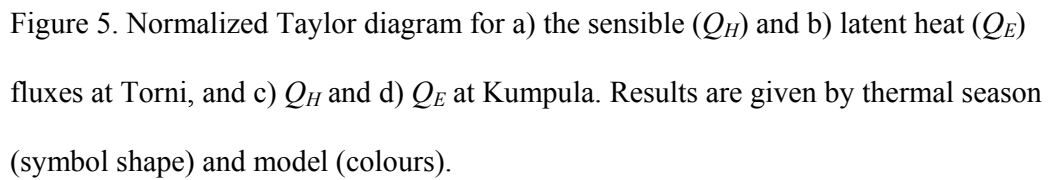


Figure 4. Modelled and observed diurnal cycles of  $Q_H$ . See Figure 3 for explanation.



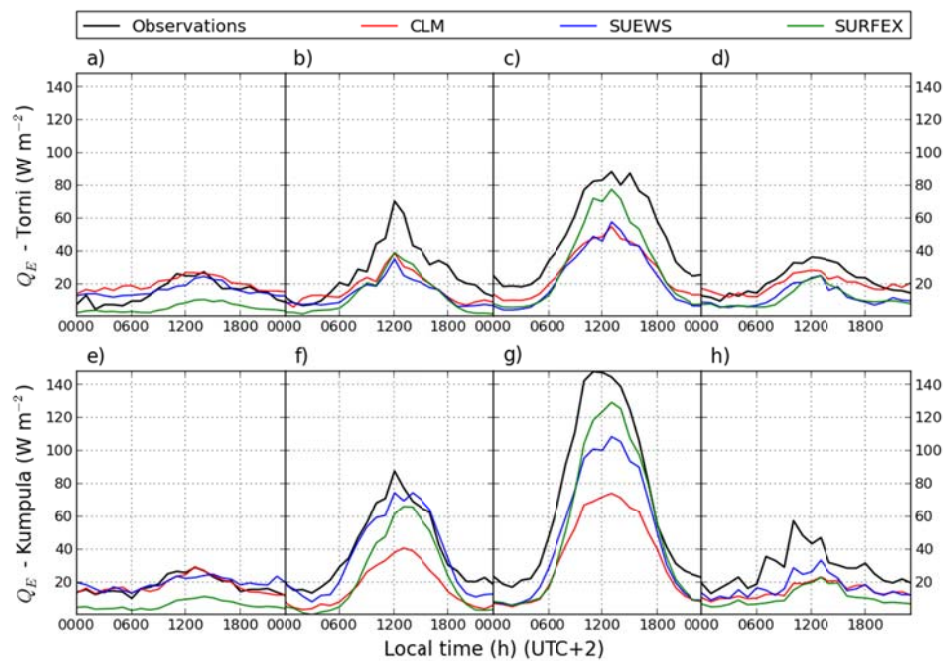


Figure 6. Modelled and observed diurnal cycles of  $Q_E$ . See Figure 3 for explanation.

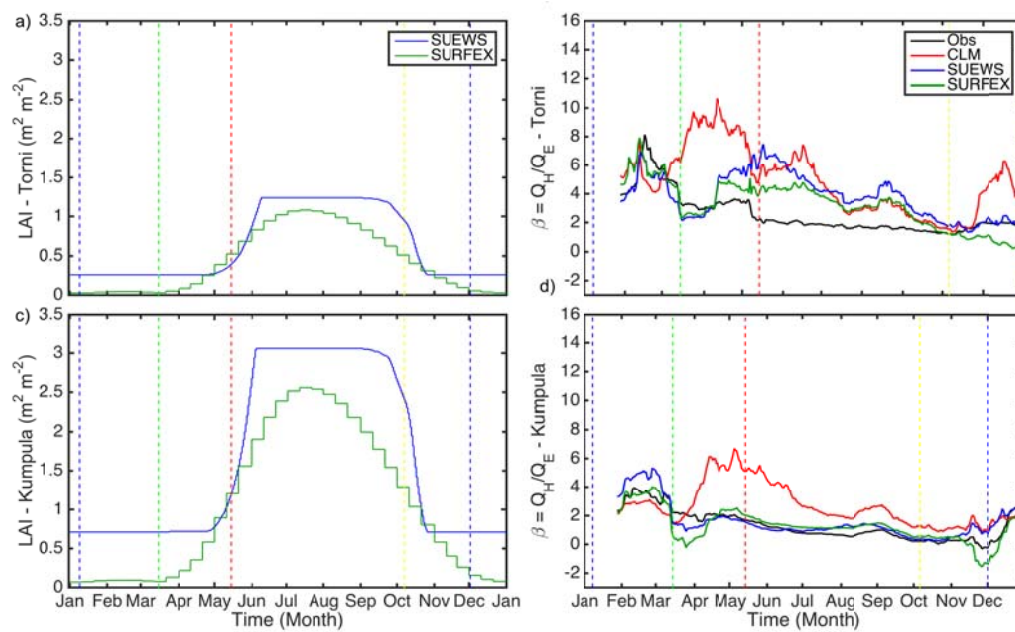


Figure 7. Behaviour of leaf area index (LAI,  $\text{m}^2 \text{m}^{-2}$ ) in SUEWS and SURFEX in a) Tornio and c) Kumpula at the vegetation sector, and 28-day running mean of Bowen ratio ( $\beta$ ) in b) Tornio and d) Kumpula in 2012. LAI is calculated as weighted average to the study areas. The dashed blue, green, red and yellow lines indicate the starting days for thermal (changes with year) winter, spring, summer and autumn, respectively.

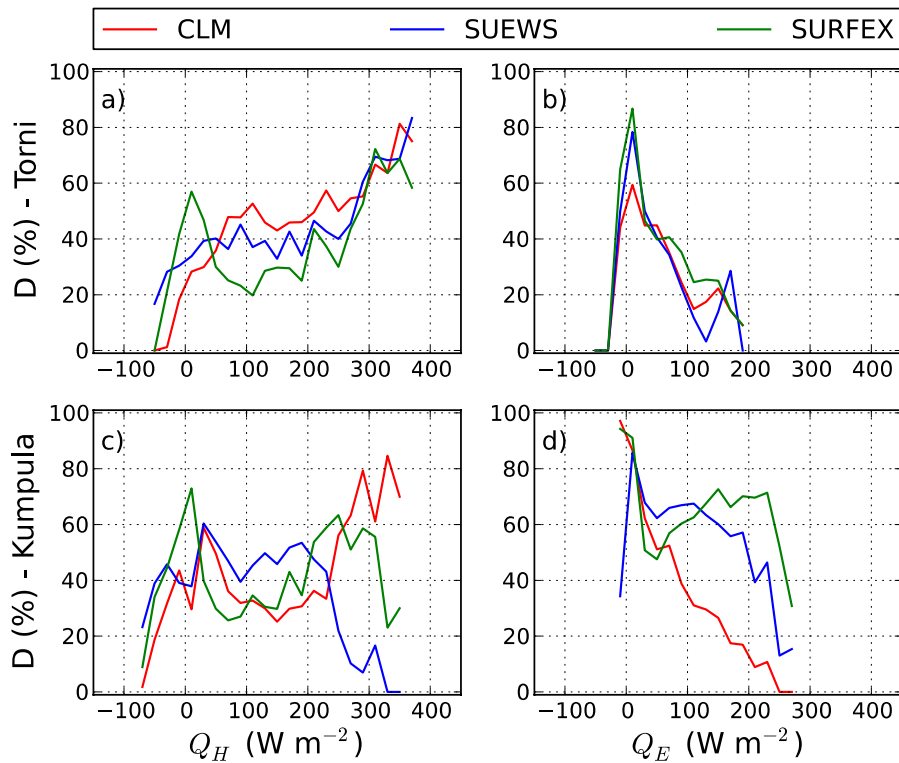


Figure 8. Normalized acceptable deviation ( $D$ ) as a function of observed turbulent flux calculated for 20  $\text{W m}^{-2}$  bins. See text for details.



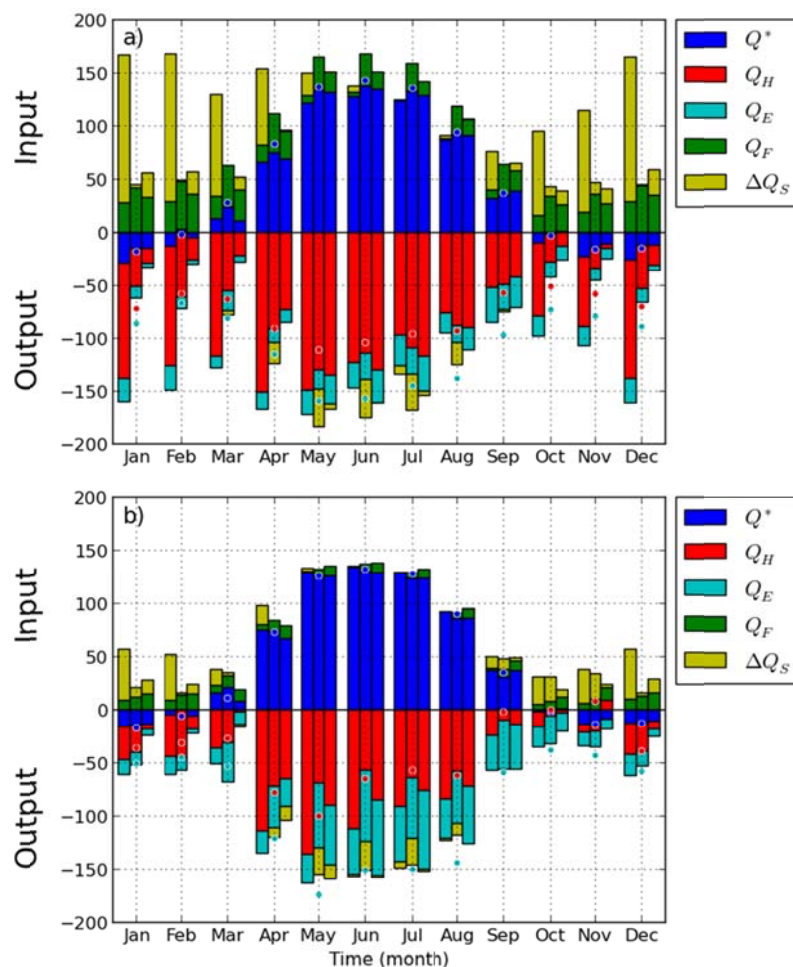


Figure 9. Monthly average energy balance (in  $\text{W m}^{-2}$ ) at a) Torni and b) Kumpula. Different fluxes are denoted with different colours. Each month has three bars: CLM (left), SUEWS (middle), SURFEX (right), and the observations of net all-wave radiation ( $Q^*$ ), sensible ( $Q_H$ ) and latent heat ( $Q_E$ ) fluxes are shown with coloured circles.

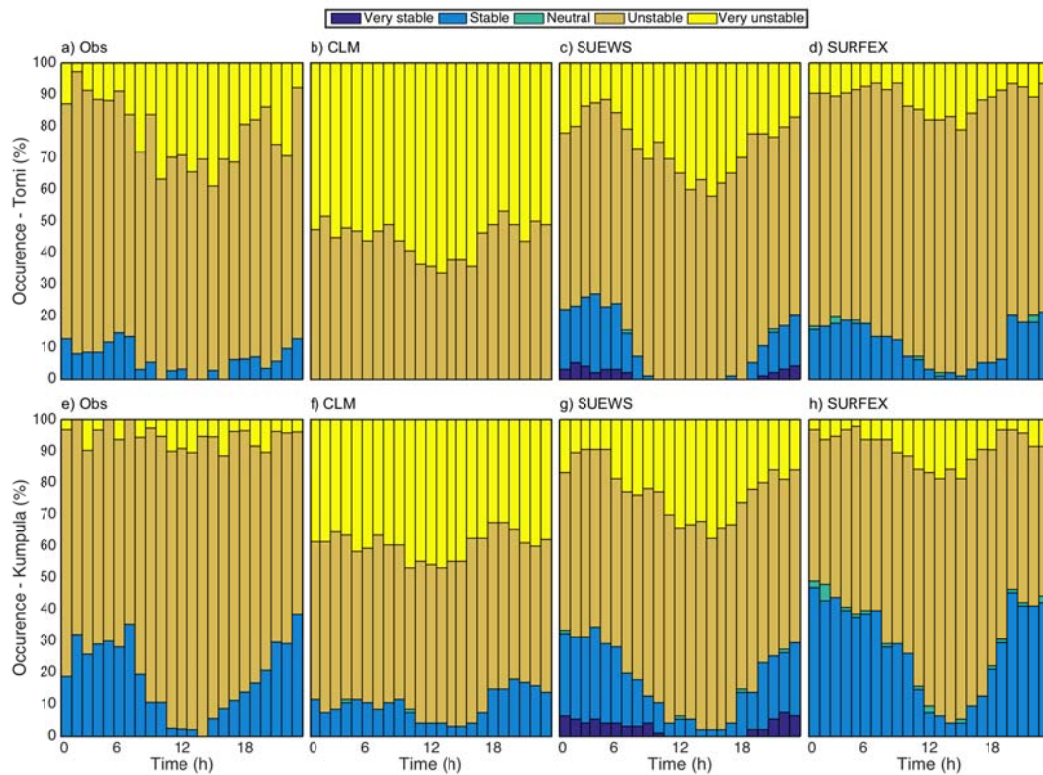


Figure 10. The occurrence of different stability classes in winter. a) observations, b) CLM, c) SUEWS and d) SURFEX at Torni, and e) observations, f) CLM, g) SUEWS and h) SURFEX at Kumpula. The different classes are (Eq. 5) very stable ( $\zeta > 0.1$ ), stable ( $0.00001 \leq \zeta \leq 0.1$ ), neutral ( $|\zeta| < 0.00001$ ), unstable ( $-0.01 \leq \zeta \leq -0.00001$ ) and very unstable ( $\zeta < -0.01$ ).